Analysis of stationary and non-stationary long memory processes : estimation, applications and forecast

Zhiping Lu

THESE DE DOCTORAT
DE L'ECOLE NORMALE SUPERIEURE DE CACHAN

Présentée par **LU Zhiping**

pour obtenir le grade de

DOCTEUR DE L'ECOLE NORMALE SUPERIEURE DE CACHAN

Domaine :
MATHEMATIQUES–MATHEMATIQUES FINANCIERES ET STATISTIQUES APPLIQUEES

Sujet de la thèse : Analyse des Processus Longue Mémoire Stationnaires et Non-stationnaires: Estimations, Applications et Prévisions

Thèse présentée et soutenue à Cachan le 02 juin 2009 devant le jury composé de :

Yves Meyer	Professeur émérite à l'Ecole Normale Supérieure de Cachan	Président
Gilles Dufrénot	Professeur à l'Université d'Aix-Marseille II	Rapporteur
Rongming WANG	Professeur à l'Ecole Normale de la Chine de l'Est	Rapporteur
Dominique Guégan	Professeur à l'Université Paris I	Directrice de Thèse
Feng ZHOU	Professeur à l'Ecole Normale de la Chine de l'Est	Directeur de Thèse

Laboratoire C.E.S. Centre d'Economie de la Sorbonne.
(ENS CACHAN UMR CNRS 8533)
61, avenue du Président Wilson - 94235 CACHAN CEDEX (France)

ECOLE NORMALE SUPERIEURE DE CACHAN
EAST CHINA NORMAL UNIVERSITY

ANALYSIS OF STATIONARY AND NON-STATIONARY LONG MEMORY PROCESSES: ESTIMATION, APPLICATIONS AND FORECAST

University:	Ecole Normale Supérieure de Cachan and East China Normal University
Department:	CES de l'Université Paris 1 and Mathematics Department
Major:	Applied Mathematics
Subject:	Mathematical Finance and Statistical Applications
Tutor:	Prof. Dominique GUEGAN and Prof. ZHOU Feng
Author:	LU Zhiping

2009. 03

Acknowledgement

This thesis would not have been possible without the encouragement and guidance of many individuals. Since I am in the cooperation program between East China Normal University and Ecole Normale Supérieure de Cachan, I would like to show my sincere thanks to many people and many organizations.

I would like to express my very deep gratitude to my French supervisor, Professor Dominique GUEGAN, who guided me at each and every aspect of this thesis. She has always listened to me patiently, sharing her insight into the applied statistics and encouraged me to solve the questions independently. She clearly answered all my questions and helped me better understand theory. I value her constructive criticisms of my work and writing. She showed me that working hard in academe can be rewarding and satisfying. I am looking forward to more collaborations in the future.

I would like to thank Professor Feng ZHOU, my Chinese supervisor, for helping me personally and professionally all these years, my graduate studies included. He taught me how to do the research, how to enlarge the scope of academic knowledge, how to live and to enrich the life in France, etc. He is always available for helping me. In particular, when I was ill, he was so anxious and have given me a lot of useful advices. I really appreciate his kindness and precious help! I could not have asked for a better mentor and professor!

I am quite grateful to Pr. Yves Meyer. He accepted to be the president of the jury, which is such a great honor for me. During the last several years, he is always very kind and has given me many precious helps. Without his signatures, I would not have registered in ENS Cachan for all these three years.

I would like to thank Professor Gilles Dufrénot. He agreed to serve as one referee of my thesis. I am quite grateful for his help.

And I wish to express my thanks to Professor Rongming WANG, for he agreed to be the other referee of the thesis and to come all the way from Shanghai to Paris for my defense.

I would like to thank Laurent Ferrara, one of my professors and my friends. He taught me a lot in the last three years. The cooperation with him is very pleasant and instructive. His kindness and encouragements gave me much motivation to do the research.

I would like to thank Pr. Dong YE and his wife, Pr. Xiaonan MA and his wife, for their friendly and warmly help and suggestion.

I also would like to thank the faculty and staff. I am very pleased to work in the laboratory of CES (Centre d'Economie de la Sorbonne) of l'University of Paris 1. I am quite grateful for the help of professors of Université Paris 1 during my stay in Paris, such as Pr. Phillipe BICH, Pr. Pascal Gourdel, Pr. Jean-Phillipe MEDECIN, Mme Marie-Lou, M. Cuong Le Van, etc. In particular, I would show my sincere thanks to Pr. Phillipe BICH. To be his teaching assistant is really pleasant and interesting. Besides, the professors of Ecole Normale Supérieure de Cachan are also very kind to me. For example, Mme Christine ROSE, the secretary of the EDSP (Ecole Doctorale de Science et de Pratique), is always patient and kind to answer my questions. Mme Vidale, the secretary of the Students' life, has helped me a lot when I applied for the "titre de séjour" in Fracne. What's more, the professors of East China Normal University are always with me when I need any help. So I would like to thank M. Wenhui YOU, M. Haisheng LI, M. Zhibin LI, Ms. Yunhua QIAN, Ms. Xiaoling LIU, Ms. Ying'e HUANG, Ms. Yujun QIN, Ms Jie XU, etc.

I would like to thank my fellow students and my friends. Qinying TAO and Beijia ZHU, the two girls have helped me a lot in the last year. I would like to show the sincerest gratitude to them for everything that they have done for me, for taking care of me when I was ill, for accompanying me to the laboratory everyday, for helping me check the first draft of the thesis, for always being so considerate, etc. I am also very grateful to M. Chunyuan ZHOU, Ms. Qiuying LU, Ms. Ting WU, Ms. Xiaoju NI, M. Hua YI, M. Tong WU, M. Haibin ZHAO, Ms. Na LI, Ms. Hua REN, M. Jie XU, M. Sanjun ZHANG, Ms. Yeqin ZHAO, M. Zhongwei TANG, M. Chun LI, Ms. Keguang CHENG, M. Chenjiang ZHU, M. Liang WANG, M. Jianxiao YANG, M. Jianxin YANG, M. Guang WU, M. Xiaopeng HE, Ms.Yaxin PENG, M. Rui HUANG, M. Lei WU, Ms. Tong LI, etc.

My deepest gratitude goes to my parents, my sister, my parents-in-law and my husband for their total and unwavering support during my studies. They taught me by their love and living example that all things are possible with hard work and discipline. Most of all, they taught me to live humble and sincere.

Thanks to everyone who has given me their valuable time, skills and enthusiasm during these last years!

Finally, I would like to show my thanks to the departments who has awarded me the scholarships in the last three years: East China Normal University (the scholarship of ECNU), Egide of France (the scholarship of Eiffel Doctorat), China Scholarship Council (the scholarship of Chinese government). Their financial aides are also important for the completion of this thesis.

LU Zhiping
2009.3

Abstract

In this thesis, we consider two classes of long memory processes: the stationary long memory processes and the non-stationary long memory processes. We are devoted to the study of their probabilistic properties, estimation methods, forecast methods and the statistical tests.

Stationary long memory processes have been extensively studied over the past decades. It has been shown that some long memory processes have the properties of self-similarity, which are important for parameter estimation. We review the self-similar properties of continuous-time and discrete-time long memory processes. We establish the propositions that stationary long memory process is asymptotically second-order self-similar, while stationary short memory process is not asymptotically second-order self-similar. Then we extend the results to specific long memory processes such as k-factor GARMA processes and k-factor GIGARCH processes. We also investigate the self-similar properties of some heteroscedastic models and the processes with switches and jumps.

We make a review for the stationary long memory processes' parameter estimation methods, including the parametric methods (for example, maximum likelihood estimation, approximate maximum likelihood estimation) and the semiparametric methods (for example, GPH method, Whittle method, Robinson method). The consistency and asymptotic normality behaviors are also investigated for the estimators.

Testing the fractionally integrated order of seasonal and non-seasonal unit roots of the stochastic stationary long memory process is quite important for the economic and financial time series modeling. The widely used Robinson test (1994) is applied to various well-known long memory models. Via Monte Carlo experiments, we study and compare the performances of this test using several sample sizes, which provide a good reference for the practitioners who want to apply Robinson's test.

In practice, seasonality and time-varying long-range dependence can often be observed and thus some kind of non-stationarity exists inside the economic and financial data sets. To take into account this kind of phenomena, we review the existing non-stationary processes and we propose a new class of non-stationary stochastic process: the locally stationary k-factor Gegenbauer process. We describe a procedure of estimating consistently the time-varying parameters with the help of the discrete wavelet packet transform (DWPT). The consistency and asymptotic normality of the estimates are proved. The robustness of the algorithm is investigated through simulation study.

We also propose the forecast method for this new non-stationary long memory processes. Applications and forecasts based on the error correction term in the error correction model of the Nikkei Stock Average 225 (NSA 225) index and the West Texas Intermediate (WTI) crude oil price are followed.

KEY WORDS : Discrete wavelet packet transform, Gegenbauer process, Long memory processes, Monte Carlo simulations, Nikkei Stock Average 225 index, Non-stationarity, Ordinary least squares estimation, Seasonality, Self-similarity, Test.

Résumé

Dans cette thèse, on considère deux types de processus longues mémoires: les processus stationnaires et non-stationnaires. Nous nous consacrons à l'étude de leurs propriétés statistiques, les méthodes d'estimation, les méthodes de prévision et les tests statistiques.

Les processus longue mémoire stationaires ont été largement étudiés au cours des dernières décennies. Il a été démontré que des processus longue mémoire ont des propriétés d'auto-similarité, qui sont importants pour l'estimation des paramètres. Nous passons en revue les propriétés d'auto-similairité des processus longue mémoire en temps continu et en temps discret. Nous proposons deux propositions montrant que les processus longue mémoire sont asymptotiquement auto-similaires du deuxième ordre, alors que processus courte mémoire ne sont pas asymptotiquement auto-similaires du deuxième ordre. Ensuite, nous étudions l'auto-similairité des processus longue mémoire spécifiques tels que les processus GARMA à k facteurs et les processus GIGARCH à k facteurs. Nous avons également étudié les propriétés d'auto-similarités des modèles heteroscedastiques et des processus avec des sauts.

Nous faisons une revue des méthodes d'estimation des paramètres des processus longue mémoire, par méthodes paramétriques (par exemple, l'estimation par maximum de vraisemblance et estimation par pseudo-maximum de vraisemblance) et les méthodes semiparamétriques (par exemple, la méthode de GPH, la méthode de Whittle, la méthode de Robinson). Les comportements de consistance et de normalité asymptotique sont également étudiés pour ces estimateurs.

Le test sur l'ordre fractionnaire intégré de la racine unité saisonnière et non-saisonnière des processus longue mémoire stationnaires est très important pour la modélisation des series économiques et financières. Le test de Robinson (1994) est largement utilisé et appliqué aux divers modèles longues mémoires bien connus. A partir de méthode de Monte Carlo, nous étudions et comparons les performances de ce test en utilisant plusieurs tailles d'échantillons. Ce travail est important pour les praticiens qui veulent utiliser le test de Robinson.

Dans la pratique, lorsqu'on traite des données financières et économiques, la saisonnalité et la dépendance qui évolvent avec le temps peuvent souvent être observées. Ainsi une sorte de non-stationnarité existe dans les données financières. Afin de prendre en compte ce genre de phénomènes, nous passons en revue les processus non-stationnaires et nous proposons une nouvelle classe de processus stochastiques: les processus de Gegenbauer à k facteurs localement stationnaire. Nous proposons une procédure d'estimation de la fonction de paramètres en utilisant la transformation discrète en paquets d'ondelettes (DWPT). La robustesse de l'algorithme est étudiée par simulations.

Nous proposons également des méthodes de prévisions pour cette nouvelle classe de processus non-stationnaire à long mémoire. Nous dennons des applications sur le terme de la correction d'erreurs de l'analyse cointégration fractionnaire de l'index Nikkei Stock

Average 225 et nous étudions les prix mondiaux du pétrole brut.

Mots Clés: Transformation discrète en paquets d'ondelettes, Processus de Gegenbauer, Processus longue mémoire, Simulation de Monte Carlo, Nikkei 225, Non-stationarité, Estimation des moindres carrés, Seasonnarité, Auto-similarité, Test.

JEL Classification: C12, C13, C14, C15, C22, C63, G15.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

In our endeavours to understand the changing world around us, observations of one kind or another are frequently made sequentially over time. The record of sunspots is a classic example, which may be traced back to 28 B.C.

A time series is a set of observations X_t, each one being recorded at a specified time t. A discrete time series is one in which the set T_0 of times at which observations are made is a discrete set, as is the case for example when observations are made at fixed time intervals. Continuous time series are obtained when observations are recorded continuously over some time interval. And the most important objective in our study of a time series is to uncover the law governing its generation.

Definition 1.0.1. *A stochastic process is a family of random variables* $\{X_t, t \in T\}$ *defined on a probability space* $(\Omega, \mathfrak{F}, P)$*, where* T *is a set of time points.*

In this thesis, we consider the data coming in the form of a univariate stochastic process. The identification of processes with strong correlation between observations far apart in time or space (so-called long memory or long range dependence) is now widespread in many diverse fields and disciplines, which is the main interest in this thesis.

The past several decades have witnessed an increasing interest in fractionally integrated processes as a convenient way of describing the long memory properties of many time series. In the pioneering papers about long memory, Granger and Joyeux (1980) and Hosking (1981) proposed a fractionally integrated ARMA (ARFIMA) process which extends the conventional ARIMA model using a differencing operator $(I - B)^d$, where d is allowed to take any non-integer value. Since then, ARFIMA processes, which have been fitted to many time series, have provided reliable long-term inferences and have become one of the most popular parametric long memory models in the literature. A singularity or a pole can be observed at the zero frequency in the spectrum of the ARFIMA process.

In empirical studies, fractional integration and long memory have been found relevant in many areas in macroeconomics and finance. There have been many studies to provide a theoretical motivation for fractional integration and long memory, for instance models based on aggregation have been suggested by Robinson (1976) and Granger and Joyeux

(1980), error duration models by Parke (1999), and regime switching models by Diebold and Inoue (2001). Other examples of applications are Diebold and Rudebusch (1989, 1991) and Sowell (1992b) for various GDP measures, Gil-Alana and Robinson (1997) for the extended Nelson-Plosser data set, Hassler and Wolters (1995) and Baillie et al. (1996) for inflation data, Diebold et al. (1991) and Baillie (1996) for real exchange rate data, and Andersen, Bollerslev, Diebold and Ebens (2001) and Andersen, Bollerslev, Diebold and Labys (2001) for financial volatility series.

One extension of the ARFIMA process is Gegenbauer ARMA (GARMA) model proposed by Gray et al. (1989) in which the operator $(I - 2\nu B + B^2)^d$ is used to characterize seasonal long memory with period $2\pi/f$ where $f(= \cos^{-1}(\nu))$ is called the Gegenbauer frequency. In comparison with the ARFIMA model, the long memory property of a GARMA process is reflected by the unbounded spectral density $f(\omega)$ around the Gegenabuer freqeuncy f:

$$f(\omega) \sim |\omega - f|^{-2d} \ as\, \omega \to f,$$

for $0 < d < 1/2$, where $x \sim y$ means that x/y tends to 1. Thus, GARMA process has only one singular point in the spectral density. The GARMA model was further generalized to allow multiple poles in the spectrum which leads to the k-factor GARMA models (see Giraitis and Leipus, 1995; Woodward et al., 1998).

Another extension was proposed by Porter-Hudak (1990), who generalized the ARFIMA models to seasonal fractionally integrated ARMA (SARFIMA) models by incorporating the operator $(I - B^s)^d$ with a known period s, with a successful application to the U.S. monetary aggregates. In contrast to the ARFIMA model, the spectral density of an SAFRIMA process has singularities not only at zero frequency but also at frequencies with multiple values of $2\pi/s$.

More generally, Robinson (1994) proposed the season and/or cyclical long memory (SCLM) model which is a generalization of most long memory processes. It permits the singularities at zero and non-zero frequencies at the same time.

In the thesis, we make great efforts on the study of the long memory processes. Briefly speaking, this thesis is composed mainly by two parts of work. The first part of our work is concerned with the modeling and probabilistic properties of stationary long memory processes. The second part of our work is concentrated on the study of non-stationary long memory processes by wavelet methods.

In the first part of our work, we review the self-similar concepts and properties for the continuous-time processes and discrete-time processes. We particularly focus on the relationship between long range dependence and self-similarity for discrete-time processes, and we present some new results for these latter models. Our major contribution here is that we have proved that if a process is both covariance stationary and long memory, then it is asymptotically second-order self-similar. Under Gaussianity, these processes are all asymptotically self-similar. We apply this result to the processes such as the k-factor

GARMA processes, the k-factor GIGARCH processes, etc. We also prove that if a process is covariance stationary and short memory, then it is not asymptotically self-similar. Thus, the usual linear ARMA processes, the GARCH processes, the models with switches and the models with breaks, are not asymptotically self-similar. Whereas, we show that under appropriate assumptions and hypotheses the SETAR models can be asymptotically self-similar. Moreover, for the processes with breaks and jumps, although they are theoretically short memory processes, they are empirically long memory processes with the sample autocorrelation function decaying hyperbolically to zero. Consequently, theoretically they are not asymptotically second-order self-similar, while empirically they are asymptotically second-order self-similar. Since we often refer the empirical long memory behavior as "spurious long memory behavior", we propose a new notion to describe this phenomenon: the "spurious asymptotically second-order self-similar" behavior. This part of work has been written as a working paper (Document de Travail du Centre d'Economie de la Sorbonne de l'Université Paris 1–2007.55) and has been presented at the seminar of "Monnaie, Banque, Finance et Assurance" of University Paris 1 in June 2007. See in Chapter 2 for more details .

Furthermore, to have an overall idea of the parameter estimation methods for the seasonal and/or cyclical long memory processes, we make a review in Chapter 4 of the parametric methods based on the maximum likelihood function in time and frequency domain, the semiparametric methods based on the Geweke and Porter-Hudak (GPH) method in the frequency domain, etc. We also discuss briefly the asymptotical behaviors of the estimators.

We are also quite interested in the statistical test for the long memory processes. Working with macroeconomics data sets significates that we deal with data sets with rather small sample sizes, generally lower than 1 000 points. Our aim is to know the accuracy of this test for a finite sample size. Our major contribution here is to assess the rate of convergence for Robinson's test using Monte Carlo simulations. To adjust these models on real data sets, it is fundamental to detect long memory behavior through statistical tests. Thus, in the second part of our work, we focus on the Robinson test (1994) to test the long memory parameters. Robinson (1994) investigated a general model in order to test whether the data stemmed from a stationary or a non-stationary process, under uncorrelated and weakly-correlated innovations $(\varepsilon_t)_t$. We recall the Robinson test and carry out the Monte Carlo simulations using grid-search method in order to study the finite sample behaviors of the test for several long memory models under the innovations of strong white noise and GARCH noise, the fractional integrated model, k-factor Gegenbauer model ($k = 1, 2, 3, 4$), rigid Hassler models, flexible Hassler models, etc. We compare the different finite sample behaviors of the seasonal and/or cyclical long memory processes, which provides a useful reference in practice. We find that the sample size is crucial for the accuracy of the test. From another point of view, the Robinson test can work as a parameter estimation method by grid searching. This part of work is described in Chapter 7 and has been presented at the seminar of "Monnaie, Banque, Finance et Assurance" of University Paris 1 in June 2008, and also at the International Symposium on Forecasting 2008 in Nice, France. This paper is in revision for the journal of Computa-

tional Statistics and Data Analysis.

Although stationarity has been usually regarded as a valid assumption for series of short duration. However, such an assumption is rapidly losing its credibility in the enormous databases maintained by firms and organizations on a large variety of subjects such as geophysics, oceanography, meteorology, speech and economics. By means of the autocorrelation function (ACF) (with autocorrelations on the y axis and the different time lags on the x axis) it is possible to detect non-stationarity of the time series with respect to the mean level.

In practice, there are many stochastic processes which are not stationary. Sometimes, we will do some transformations to make them stationary. An example of non-stationary time series is given in Figure 1.1, which represents the values of the Nasdaq-100 index from January 4, 1999 through January 4, 2009. This index includes hundred of the largest non-financial domestic and international companies listed on the Nasdaq National Market. It is clear that this process contains a trend, which can be removed if we compute the first-order difference of the logarithm of the Nasdaq-100 series (the log-return index), see Figure 1.2. The resulting zero-mean process still contains valuable information for the analyst, in terms of the volatility of the time series.

However, some time series are not suitable to be made stationary. An example of a non-stationary time series is a record of readings of the atmosphere temperature measured each 10 seconds with some random errors that have a constant distribution with zero mean. At any given time point the mean of the readings is equal to the true temperature. On the other hand, the mean value itself changes with time–as far as the true temperature varies with time. Thus, it gives us great motivations to investigate the estimation method of the long memory parameters for non-stationary processes.

To our knowledge, the existing non-stationarity of long memory time series is characterized by two kinds of parameters. One is the constant parameter $d(\in (\frac{1}{2}, 1))$ in the fractionally integrated model $(I-B)^d X_t = \varepsilon_t$. The other is the time-varying parameter function $d(t)(\in (-\frac{1}{2}, \frac{1}{2}))$ in the generalized fractionally integrated model $(I-B)^{d(t)} X_t = \varepsilon_t$ considered by Jensen (2000), Whithcer and Jensen (2000), Cavanaugh et al. (2002), etc.

It is necessary to extend the concept of long memory to the non-stationary framework, (see Cheung and Lai, 1993; Maynard and Phillips, 2001; Phillips, 2005). In Chapter 5, for non-stationary processes, we first review the previous work focusing on the estimation method of long memory parameter, in particularly the fractional integrated model. The methods are concentrated on the GPH method, local Whittle method, exact local Whittle method, fully extended local Whittle method, Whittle pseudo maximum likelihood method and wavelet-based local Whittle method. We also discuss the corresponding consistency and normality of the estimators. Hurvich and Ray (1995) argued, that the GPH estimator was consistent only when $d < 1$ by simulation, and Kim and Phillips (1999) theoretically. In the same context, Velasco (1999a) showed the consistency and

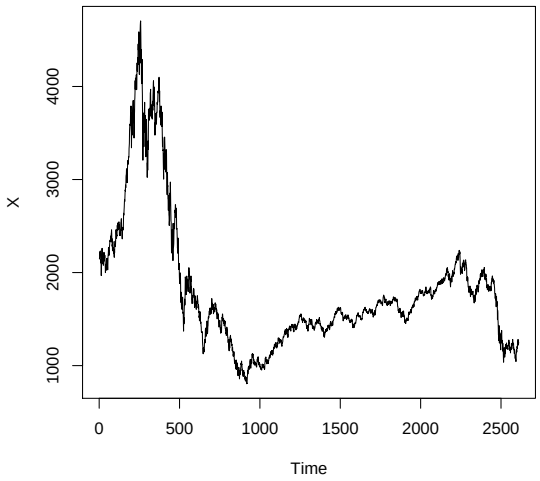


Figure 1.1: 604 daily observations of the Nasdaq-100 index

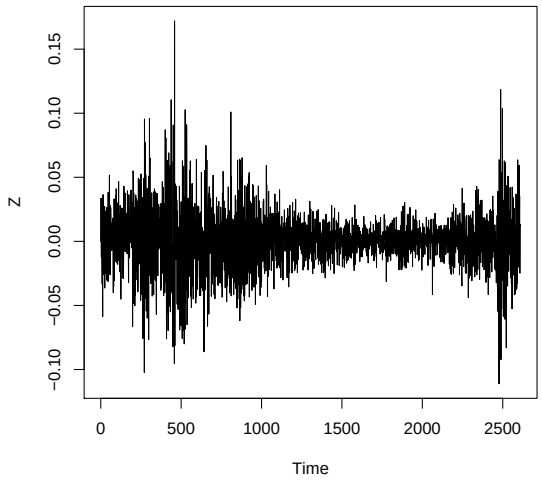


Figure 1.2: The log-returns of the Nasdaq-100 index

the asymptotic normality of the Robinson (1995a) estimator for $d \in [1/2, 3/4]$. To solve the non-consistency problem, Hurvich and Ray (1995) and Velasco (1999a) suggested the use of data tapering, which was first proposed by Cooley and Tukey (1965) and discussed by Cooley et al. (1967) and Jones (1971). This technique has also been used by many authors, such as Hurvich and Chen (2000), Giraitis and Robinson (2003), Sibbertsen (2004), Olhede et al. (2004), among many others. For any value of d, Velasco (1999a) showed that if the tapering order p is greater or equal to $[d + 1/2] + 1$, the estimator is consistent and asymptotically normal.

On the other hand, in reality, many naturally occurring phenomena show a slow drift in their periodicities and self-similarities, when observed for enough long time period. The changes in the cycles and self-similarities are often so gradual that a process over a short span will behave as a stationary process. This information on the drift of the cycles and self-similarities can provide a great deal of insight into the dynamics of an evolving process. The analysis of the non-stationary economic time series with time-varying parameters has been going on for some time. According to the long memory processes with time-varying parameter functions, one important model is the locally stationary ARFIMA model, which has been studied by Jensen (1992a), Whitcher and Jensen (2000) and Cavanaugh et al. (2002), for instance. In fact, this model can be regarded as an extension of the stationary ARFIMA model which permits the long memory parameter to evolve with time. These authors all used the wavelet techniques based on the semiparametric method for estimating the parameters. Whitcher and Jensen (2000) improved the wavelet-based GPH method (Jensen, 1999) by the concept of the cone of influence. Cavanaugh et al. (2002) studied the self-similar property of the locally stationary ARFIMA model, built the approximate log-linear relationship, and estimated the time-varying parameters which has been proved to be quite efficient. Concerning the estimation methods of this class

of processes, Brock (2000) discussed the need for time-frequency analysis of economic time series when he brought up the work on the interface between ecology and economics.

Fourier transform is known to be suitable to analyze the stationary time series with its localized frequency basis function. Whereas, since wavelet transform is localized both in time and in frequency, it has the ability to capture the characteristics of the events that are local in time, which makes it an ideal tool for studying non-stationary or transient time series with time-varying characteristics. Yves Meyer may be considered as the founder of this mathematical subject, which we call wavelet analysis. Of course, Meyer's profound contribution to wavelet analysis is much more than being a pioneer of this new mathematical field. For the past ten years, he has been totally committed to its development, not only by building the mathematical foundation, but also by actively promoting the field as an interdisciplinary area of research. In fact, a number of concepts such as non-stationarity, multiresolution and approximate decorrelation emerge from wavelet filters, as it provides a natural platform to deal with the time-varying characteristics found in most real world time series, and thus the assumption of stationarity may be avoided. Hence, a transform that decomposes a process into different time horizons (scales) is appealing as it differences seasonalities, reveals structural breaks and volatility clusters, and identifies local and global dynamic properties of a process at these time scales. Moreover, wavelet filters provide a convenient way of dissolving the correlation structure of a process across time scales. This will indicate that the wavelet coefficients at one level are not (much) associated with coefficients at different scales or within their scales. This is useful when performing tasks such as simulations, estimations, and test since it is always easier to deal with an uncorrelated process than one with unknown correlation structure.

One of the most important task of this thesis is to study the non-stationary time series models with seasonalities which permit the long memory parameters to evolve with time, motivated by the idea of Dalhaus (1997). As an extension of the k-factor GARMA model and that of the fractionally integrated model, a class of new model (locally stationary k-factor Gegenbauer process) is proposed:

$$\prod_{i=1}^{k}(I - 2\nu B + B^2)^{d(t)} y_t = \varepsilon_t,$$

where ε_t is Gaussian white noise. Fr appropriate values of $d(t)$, this model can be regarded locally as stationary processes with constant parameter, that is to say, piecewise stationary. Based on this kind of understanding, we develop a wavelet-based estimation procedure with the help of the DWPT. After evenly partitioning the time interval, we locate the DWPT coefficients on each subinterval according to the "Heisenberg box" principle. Then we calculate the the wavelet variances on each subinterval, which is an approximate estimate of the spectrum. Then we apply the ordinary least squares regressions on the subintervals. What we obtain at last are the local estimates on the subintervals. The first and last estimates are omitted before doing smoothing in order to avoid the boundary effects. We adopt the spline smoothing method and loess smoothing method to get the smoothed curves. Furthermore, we prove the consistency and asymptotical normality.

The robustness of the estimates is verified by Monte Carlo simulation study. We simulate six kinds of elementary functions as the parametric functions: the constant, linear, quadratic, cubic, exponential and logarithmic functions. We present the smoothed curves for the estimated long memory parameter functions with comparison to the other authors' estimation results, and we also present the mean bias and the mean RMSE of the estimation for the simulated series. To our knowledge, there exists almost no literature of the forecast method for the non-stationary processes. For the new non-stationary processes that we propose in this thesis, we develop the corresponding forecast method based on the estimations. According to the model, the most critical point is to make the forecast for the long memory parameter function, with which we obtain the forecast for the original series.

This part of work is presented in Chapter 5 and we have written as a working paper (Document de Travail du Centre d'Economie de la Sorbonne de l'Université Paris 1–2009.15). It will be submitted to a statistic journal of high level.

In the application part, we apply the new non-stationary model on the error correction term of the cointegration model of Nikkei Stock Average 225 index and on the WTI oil price data, using the wavelet-based estimation procedure that we proposed and we make the h-step-ahead forecasts of the long memory parameter functions. Then we make the h-step-ahead forecasts of the error correction terms. The forecasts are also evaluated by the bias and the root mean square error. We discuss a little the applicability of the new model and the new estimation method.

The remainder of this thesis is organized as follows: In Chapter 2, we introduce some probabilistic properties of stationary processes. Chapter 3 are the wavelet techniques for time series analysis. In Chapter 4, we make a review of the estimation methods for stationary long memory processes. Chapter 5 is concerned with the estimation methods of the non-stationary processes. In Chapter 6, we make some applications on the financial data series and the energy data series. Chapter 7 is concentrated on the tests for the fractional order of long memory processes. Chapter 8 is the conclusion.

Chapter 2

Some Probabilistic Properties of Stationary Processes

2.1 Introduction of Stationary Processes

Let X_t denote a real-valued random variable representing the observation made at time t. We confine our study to observations made at regular time intervals, and without loss of generality, we assume that the basic time interval is of duration one unit of time.

Now we present some important features of time series with time-invariant properties.

Definition 2.1.1. *The time series $(X_t)_t$ is said to be stricly stationary if, for any $t_1, t_2, \cdots, t_n \in \mathbb{Z}$, any $k \in \mathbb{Z}$, and $n = 1, 2, \cdots$,*

$$F_{X_{t_1}, X_{t_2}, \cdots, X_{t_n}}(x_1, \cdots, x_n) = F_{X_{t_1+k}, X_{t_2+k}, \cdots, X_{t_n+k}}(x_1, \cdots, x_n), \tag{2.1}$$

where F denotes the distribution function of the set of random variables which appear as suffices.

The term "weakly stationary", "second-order stationary", "covariance stationary" or "wide-sense stationary" is used to describe the theoretically less restricted situation in which

$$E(X_{t_1}) = E(X_{t_1+k}), \qquad Cov(X_{t_1}, X_{t_2}) = Cov(X_{t_1+k}, X_{t_2+k}), \tag{2.2}$$

for all $t_1, t_2, k \in \mathbb{Z}$, the covariances being assumed to exist. Strict stationarity implies weak stationarity provided that $Var(X_t)$ exists. And these two concepts are equivalent under Gaussianity.

In practice, to obtain stationarity, sometimes we need to do some simple transformations, such as taking differences of consecutive observations, subtracting a polynomial or trigonometric trend, etc.

Consider a covariance stationary time series $(X_t)_t$ with finite variance. It follows from the relationship (2.2) that $Cov(X_{t_1}, X_{t_2})$ is simply a function of $|t_1 - t_2|$. This function is called the autocovariance function of $(X_t)_t$ at $lag\ (t_1 - t_2)$. We denote it by $\gamma_{t_2 - t_1}$. The

ratio γ_τ/γ_0 ($\tau \in \mathbb{Z}$), is called the autocorrelation function of $(X_t)_t$ of $lag\ \tau$, denoted by ρ_τ.

For a scalar covariance stationary process $(X_t)_t$, if we assume absolute continuity of the spectral distribution function, then there is a spectral density $f(\lambda)$ ($-\pi < \lambda \leq \pi$), such that the autocovariance

$$\gamma_j = E[(X_1 - E(X_1))(X_{1+j} - E(X_1))] = \int_{-\pi}^{\pi} \cos(j\lambda) f(\lambda) d\lambda.$$

In the analysis of stationary time series, the behavior of the spectral density around zero frequency is often of interest, which is an important characteristic in identifying the model.

2.1.1 Short Memory Processes

Definition 2.1.2. *A covariance stationary process $(X_t)_t$ is called short memory (or short range dependence) process, if its autocorrelation function ρ_k satisfies $\sum_{k=0}^{\infty} \rho_k < \infty$.*

Thus the autoregressive moving average(ARMA) processes and autoregressive conditional heteroscedastic(ARCH) processes are the classical short memory processes. In Figure 2.1 we observe the sample autocorrelation function of a simulated short memory process.

2.1.2 Long Memory Processes

One of the earliest studies which mentioned that the observed time series may exhibit long range dependence is attributed to the pioneering work of Hurst (1951). While looking at time series from the physical sciences (rainfall, tress rings, river levels, etc.), he noticed that his R/S- statistic, on a logarithmic scale, was randomly scattered around a line with slope $H > \frac{1}{2}$ for large sample sizes. For a stationary process with short range dependence, $\log R/S$ should be proportional to $k^{1/2}$, for large k. Hurst's discovery of slopes proportional to k^H, with $H > 1/2$, was in direct contradiction to the theory of such processes at the time. This discovery is known as the "Hurst effect" and H is the "Hurst parameter".

Mandelbrot and co-workers (Mandelbrot and Van Ness, 1968; Mandelbrot and Wallis, 1969) showed that the Hurst effect may be modeled by fractional Gaussian noise with self-similarity parameter $0 < H < 1$. This process exhibits stationary long memory dynamics when $1/2 < H < 1$ and reduces to white noise when $H = \frac{1}{2}$. The spectrum of fractional Gaussian noise may be derived from an appropriately filtered spectral density function of fractional Brownian motion and evaluated via numeric integration, although finite term approximations may be used in practice. We will instead focus our attention on a convenient class of time series models known as fractional differencing processes.

Since the work of Granger and Joyeux (1980) and Hosking (1981), long memory processes have been widely studied. There exist different characterizations for the concept of long memory in time domain or in spectral domain.

We say that a function h changes slowly to infinity (or to zero) if it satisfies the condition **H**: for all $a \in \mathbb{R}$, $h(ax)/h(x) \to 1$, when $x \to \infty$ (or $x \to 0$).

Definition 2.1.3. *A covariance stationary process $(X_t)_t$ is called long memory process if it has an autocorrelation function ρ_k which behaves like a power function decaying to zero hyperbolically as*

$$\rho_k \sim C_\rho(k) \cdot k^{-\alpha}, \; as \;\; k \to \infty, \; 0 < \alpha < 1, \tag{2.3}$$

*where $\sim$ represents the asymptotic equivalence, and $C_\rho(k)$ is a function which changes slowly to infinity satisfying the condition **H**.*

This definition is concerned with the concept of long memory behavior in time domain, which means that the decay rate of autocorrelations is very slow. Thus, the autocorrelation series is absolutely divergent, i.e. $\Sigma_{k=0}^{\infty}\rho_k = \infty$. Remark that if we note $H = 1 - \frac{\alpha}{2}$, the long memory behavior following the Definition 2.1.3 occurs when $\frac{1}{2} < H < 1$. Figure 2.2 shows the sample autocorrelation function (ACF) of a simulated long memory process. Obviously, a very slow decay of the ACF is displayed.

Due to the Fourier transform, the covariance function is connected to the spectral density. In spectral domain, we define the long memory behavior which is characterized by the rate of explosions at low frequencies as follows:

Definition 2.1.4. *A covariance stationary process $(X_t)_t$ is called long memory process if its spectral density function f is approximated by*

$$f(\lambda) \sim C_f(\lambda) \cdot \lambda^{-2d}, \;\; as \;\; \lambda \to 0^+, \; 0 < d < \frac{1}{2}, \tag{2.4}$$

*where $C_f(\lambda)$ is a function which changes slowly to zero at frequency zero satisfying the condition **H**.*

For $d \geq \frac{1}{2}$, a function behaving like λ^{-2d} as $\lambda \to 0^+$ will not be integrable so that covariance stationarity cannot be obtained, while the case when $d > -\frac{1}{2}$ corresponds to an invertibility condition in parametric models with property (2.4). In fact, property (2.4) is also useful in modeling stochastic processes by semi-parametric methods. Figure 2.3 illustrates the spectrum of a a simulated long memory behavior. We can find an explosion at the zero frequency in the spectrum.

It should be pointed out that Definition 2.1.3 and Definition 2.1.4 are given in an asymptotic context near the zero frequency, which indicate the singularity at zero frequency in

the spectral density.

More generally, a process is said to exhibit long memory behavior if its spectral density is unbounded at a finite number of frequencies between $[0, \pi]$. While if the process exhibits short memory behavior, the spectral density is bounded on the frequency interval $[0, \pi]$.

Definition 2.1.5. *A covariance stationary process $(X_t)_t$ is called long memory process if its spectral density function f has the following property:*

$$\exists \lambda_0 \in [0, \pi], \quad such\ that\ f(\lambda_0)\ is\ unbounded.$$

Figure 2.4 shows the spectrum of a simulated long memory process (Gegenbauer process) with an explosion at the frequency $\lambda_0 = \frac{2\pi}{3}$. Figure 2.5 shows the spectrum of the 2-factor Gegenbauer process with two explosions in the spectrum. While in Figure 2.6, we can observe three explosions in the spectrum of the Hassler model.

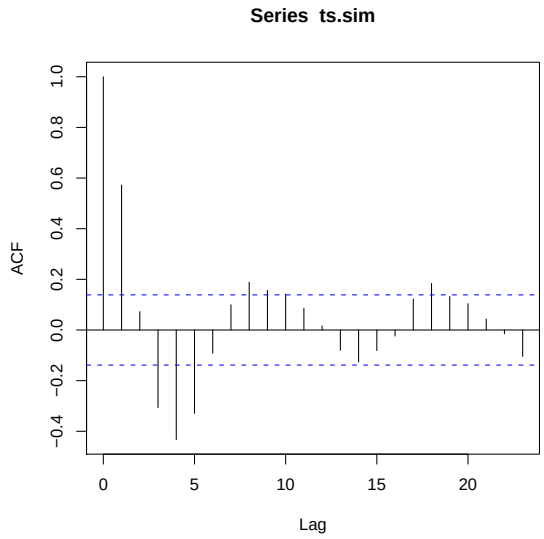


Figure 2.1: ACF of a simulated short memory process

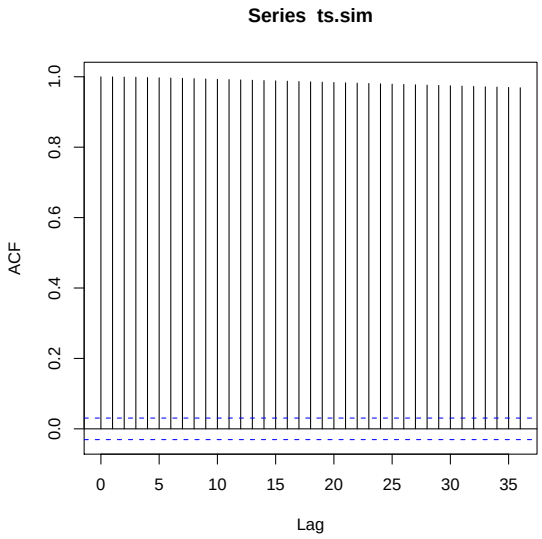


Figure 2.2: ACF of a simulated long memory process

2.2 Self-similar Properties for Stationary Processes

2.2.1 Concepts of Self-similarity

In the past decades there has been a growing interest in studying self-similar processes and asymptotically self-similar processes which were first introduced theoretically through the work of Komogorov (1941). These processes are typically used to model random phenomena with long range dependence, which informally means significant correlations across arbitrarily large time scales.

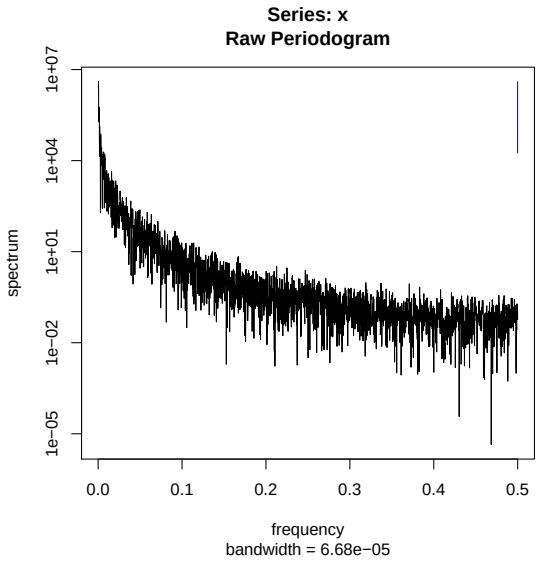


Figure 2.3: Spectrum of a simulated long memory process (FI(d) Process)

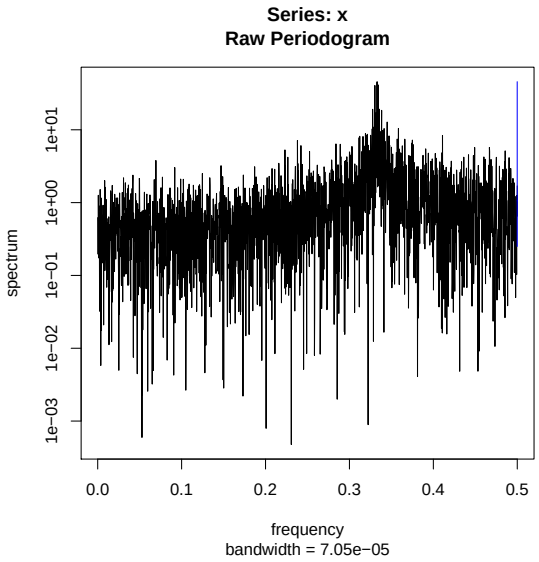


Figure 2.4: Spectrum of a simulated long memory process (Gegenbauer Process)

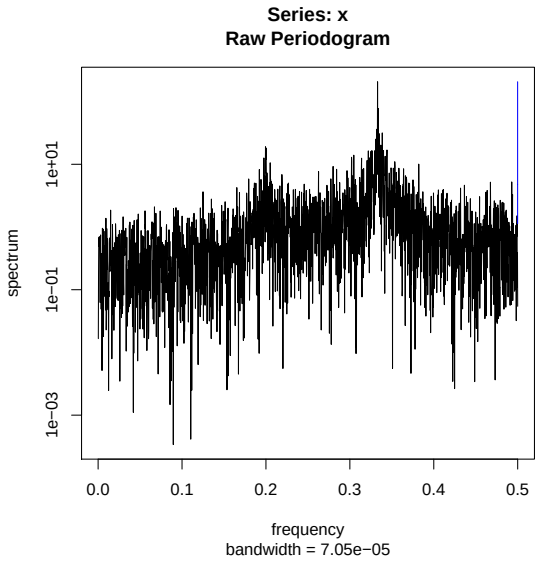


Figure 2.5: Spectrum of a simulated long memory process (2-factor Gegenbauer process)

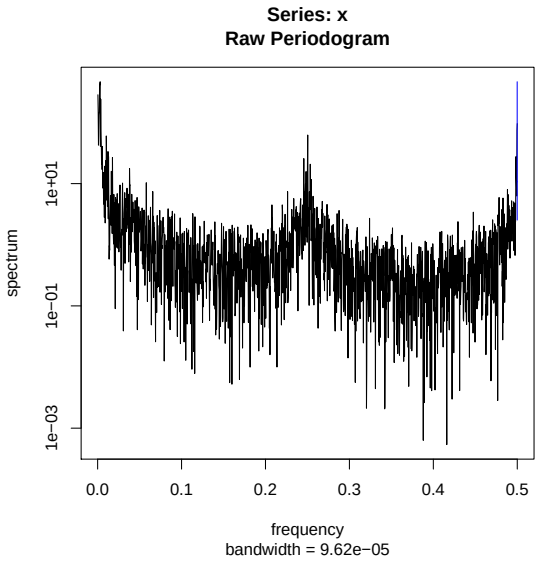


Figure 2.6: Spectrum of a simulated long memory process (Hassler Model)

The concept of self-similarity has been considered by many authors, mainly in the context of continuous time. Many applications can be found in physics, geophysics, hydrology, finance, economics, communications and the "1/f noises".

Self-similarity is a pervasive characteristic to describe the phenomena exhibiting certain forms of long range dependence. It provides an elegant explanation and interpretation of an empirical law that is commonly referred as the Hurst's effect. The parameter H introduced previously is the so-called Hurst parameter, designed to capture the degree of self-similarity. Self-similarity measures the deviation of self-similar process away from the Brownian motion whose Hurst parameter H is equal to $\frac{1}{2}$.

A priori, there is no evident link between the notion of self-similarity and long memory (see Beran (1994)). Generally speaking, the concept of long memory is not equivalent to that of self-similarity or asymptotical self-similarity. For a self-similar or asymptotically self-similar process, we cannot justify directly whether it is long memory or not. For example, Brownian motion is self-similar but not long memory, while fractional Brownian motion is an H-self-similar process with stationary increments (for specific values of H) and at the same time it exhibits long memory behavior. Thus we observe that the relationship between these two properties necessitates to be carefully identified.

There exist several different definitions for self-similarity and asymptotic self-similarity. In the context of stochastic processes, self-similarity is defined in terms of the distribution of the processes. We will consider these concepts respectively for continuous-time and discrete-time processes.

2.2.2 Continuous-time Self-similar Processes

Definitions of Self-similarity

Definition 2.2.1. *A real-valued stochastic process $Z = \{Z(t)\}_{t\in\mathbb{R}}$ is an H-self-similar process (or self-similar process with index $H > 0$), if for any $a > 0$,*

$$\{Z(at)\}_{t\in\mathbb{R}} =^d \{a^H Z(t)\}. \tag{2.5}$$

We call such a model H-ss process.
Here " $=^d$ " means identical in finite-dimensional distributions.

Definition 2.2.2. *A process $Z = \{Z(t)\}_{t\in\mathbb{R}}$ is an H-self-similar process with stationary increments (or self-similar process with index $H > 0$ and with stationary increments), if for the H-ss process $\{Z_t\}_{t\in\mathbb{R}}$ and all $h \in \mathbb{Z}$,*

$$\{Z(t+h) - Z(h)\}_{t\in\mathbb{R}} =^d \{Z(t) - Z(0)\}_{t\in\mathbb{R}}. \tag{2.6}$$

We call such a model H-sssi process.

Properties of Self-similar Processes

Let $\{Z_t\}_{t\in\mathbb{R}}$ be an H-sssi process with $0 \leq H < 1$. We introduce the increment process $(X_t)_{t\in\mathbb{Z}}$ defined by $X_t = Z_t - Z_{t-1}$, $\forall\ t$. Denote $\gamma_Z(\cdot)$ the covariance function of the process $\{Z_t\}_{t\in\mathbb{R}}$ and σ_Z^2 its variance. Then we have the following properties, see Beran (1994) and Taqqu et al. (1997):

1. Z(0)=0 almost surely (a.s.).
2. $-Z(t) =^d Z(-t)$.
3. Suppose $Z(t)$ is a non-trivial H-sssi process with $H > 0$:
   (i) If $E[|Z(1)|^\gamma] < \infty$ for some $\gamma < 1$, then $H < 1/\gamma$;
   (ii) If $E[|Z(1)|] < \infty$, then $H < 1$;
   (iii) If $E[|Z(1)|] < \infty$ and $0 < H < 1$, then $E[Z(t)] = 0$;
   (iv) If $E[|Z(1)|] < \infty$ and $H = 1$, then $Z(t) = tZ(1)\ a.s.$.
4. $E[(Z_t - Z_{t-1})^2] = \sigma_Z^2(t-s)^{2H}$ and $\gamma_Z(t,s) = \dfrac{1}{2}\sigma_Z^2[t^{2H} - (t-s)^{2H} + s^{2H}]$.
5. Let $\overline{X}$ be the sample mean of the process $(X_t)_t$, then we have some empirical properties concerning $\overline{X}$: $\overline{X} = n^{H-1}(Z_1 - Z_0)$ and $Var(\overline{X}) = \sigma_Z^2 n^{2H-2}$.
6. If $(X_t)_t$ is a Gaussian process with mean μ and variance σ_X^2, then $n^{1-H}(\overline{X} - \mu)/\sigma_X$ is a standard normal random variable.
7. Among continuous-time processes, the H-ss process and the H-sssi process are the most frequently studied processes. Definition 2.2.1 hints that self-similar processes are stochastic processes that are invariant in distribution under appropriate scaling of time and space. If an H-ss process is not covariance stationary, we will consider the increments of the H-ss process to get some stationarity. On the other hand, an H-ss process cannot be strictly stationary.
8. For the $H-ss$ process $\{Z(t)\}_{t\in\mathbb{R}}$, the process $Y(t) = e^{-tH}Z(e^t)(t \in \mathbb{R})$ is stationary. Conversely, if $\{Y(t)\}_{t\in\mathbb{R}}$ is stationary, then the process $Z(t) = t^H Y(lnt)(t > 0)$ is H-ss.
9. When $\dfrac{1}{2} < H < 1$, the process $(X_t)_{t\in\mathbb{Z}}$ exhibits long memory behavior.
   When $H = \dfrac{1}{2}$, the process $(X_t)_{t\in\mathbb{Z}}$ is uncorrelated.
   When $0 < H < \dfrac{1}{2}$, the process $(X_t)_{t\in\mathbb{Z}}$ exhibits short memory.

2.2.3 Discrete-time Self-similar Processes

For a covariance stationary time series $(X_t)_{t\in\mathbb{Z}}$, if its covariance exists and its autocorrelation function ρ_k has the asymptotical behavior $\lim_{k\to\infty} \rho_k = 0$, then the parameter H belongs to the interval $(0, 1)$.

Let $(X_t)_t$ be a covariance stationary time series. Denote

$$X_t^{(m)} = \frac{1}{m} \sum_{k=(t-1)m+1}^{tm} X_k, \qquad k = 1, 2, ..., \tag{2.7}$$

the corresponding aggregated sequence with level of aggregation m(> 1), and $\rho_k^{(m)}$ the autocorrelation function of the process $(X_t^{(m)})_{t\in\mathbb{Z}}$.

Definitions of Self-similarity

Definition 2.2.3. *A strictly stationary stochastic process $(X_t)_t$ is exactly self-similar (or asymptotically self-similar) if for all t,*

$$X_t =^d m^{1-H} X_t^{(m)} \tag{2.8}$$

holds for all m (or as $m \to \infty$), where $(X_t^{(m)})_t$ is the aggregated sequence defined as in Equation (2.7) *and $1/2 < H < 1$.*

Another characterization of self-similarity is based on second-order moments as follows:

Definition 2.2.4. *Let $(X_t)_t$ be a covariance stationary process,*

1. *The process $(X_t)_t$ is called exactly second-order self-similar, or s.o.s.s, if $m^{1-H} X_t^{(m)}$ has the same autocorrelation as X_t, for all m and for all t. Thus,*

$$Var(X^{(m)}) = Var(X) m^{2H-2} \tag{2.9}$$

$$and \;\; \rho_k^{(m)} = \rho_k \tag{2.10}$$

where $1/2 < H < 1$, $m > 1$, $k = 0, 1, 2, \cdots$, and $\rho_k \sim Ck^{2H-2}$, $as\ k \to \infty$.

2. *The process $(X_t)_t$ is called asymptotically second-order self-similar, or a.s.o.s.s, if*

$$\lim_{m\to\infty} \rho_k^{(m)} = \frac{1}{2}[(k+1)^{2H} - 2k^{2H} + (k-1)^{2H}],\ \forall\ k > 0.$$

We denote in the following

$$g^H(k) = \frac{1}{2}[(k+1)^{2H} - 2k^{2H} + (k-1)^{2H}], \;\; where\ k > 0. \tag{2.11}$$

The notion of exact (asymptotical) self-similarity concerns all the finite-dimensional distributions of a strictly stationary process, while the notion of exact (asymptotical) second-order self-similarity concerns only the variance and autocorrelation function of a covariance stationary process. In fact, under Gaussian framework, exact second-order self-similarity is equivalent to exact self-similarity, and likewise, asymptotic second-order self-similarity is equivalent to asymptotic self-similarity.

Properties of s.o.s.s. and a.s.o.s.s Time Series

Let $(X_t)_t$ be a covariance stationary exactly second-order self-similar process. Denote ρ_k its autocorrelation function, $f_X(\lambda)$ its spectral density function, σ_X^2 its variance, then we have the following properties:

1. The autocorrelation function of the s.o.s.s process $(X_t)_t$ is such that:

$$\rho_k = g^H(k), \quad \forall k > 0, \tag{2.12}$$

   where $g^H(k)$ is introduced in the Equation (2.11).

2. For $1/2 < H < 1$, the s.o.s.s process $(X_t)_t$ exhibits long memory behavior with its autocorrelation function decaying hyperbolically: $\rho_k \sim H(2H-1)k^{2H-2}$, as $k \to \infty$. And, its spectral density function explodes at the origin with $f_X(\lambda) \sim |\lambda|^{1-2H}$, as $\lambda \to 0$.

3. The spectral density of the process $(X_t)_t$ verifies:

$$f_X(\lambda) = 2C_f(1-\cos\lambda)\sum_{i=-\infty}^{\infty} \mid 2\pi i+\lambda \mid^{-2H-1} = C_f \mid \lambda \mid^{1-2H} + O(\mid \lambda \mid^{min(3-2H,2)}) \tag{2.13}$$

   where $\lambda \in [-\pi, \pi]$, $1/2 < H < 1$ and $C_f = \sigma_X^2(2\pi)^{-1}\sin(\pi H)\Gamma(2H+1)$.

4. There exists some equivalences between the previous properties. The condition (2.9) is equivalent to the condition (2.12) and they are both equivalent to the condition (2.13). Whereas, with the condition (2.9), we can deduce condition (2.10), but not vice versa. (see Taqqu et al. (1997) for more details)

It should be pointed out that we can also use the above property 1 and property 3 as the definition of second-order self-similarity processes. Thus, with these properties, we obtain the following result which presents the relationship between self-similarity and long memory behaviors.

Lemma 2.2.5. *Let $(X_t)_t$ be a covariance stationary process, if this process is short memory as is defined in Definition 2.1.2, then it is not asymptotically second-order self-similar.*

Proof. For a short memory process $(X_t)_t$, the corresponding autocorrelation function decays exponentially to zero, so it does not satisfy the Equation (2.12), which means that the process $(X_t)_t$ is not asymptotically second-order self-similar. □

Lemma 2.2.6. *Let $(X_t)_t$ be a covariance stationary long memory process with $\frac{1}{2} < H < 1$, then this process is asymptotically second-order self-similar. Furthermore, under Gaussianity, the process is asymptotically self-similar.*

Proof. For a covariance stationary process $(X_t)_t$, its autocorrelation function decays hyperbolically, i.e.

$$\lim_{k\to\infty} \frac{\rho_k}{k^{2H-2}} = c, \ \frac{1}{2} < H < 1.$$

According to the results of Tsybakov and Georganas (1997), we deduce that

$$\lim_{m\to\infty} \rho_k^{(m)} = g^H(k),\ \forall k = 1, 2, \cdots .$$

Thus the process is asymptotically second-order self-similar following the definition 2.2.4. □

From the Lemma 2.2.6, the following result is straightforward:

Lemma 2.2.7. *Let $(X_t)_t$ be a covariance stationary long memory process with $\frac{1}{2} < H < 1$, if the spectral density of the process blows up at the origin, then the process is asymptotically second-order self-similar. Under Gaussianity, it is asymptotically self-similar.*

2.2.4 Examples of Self-similar Processes in Continuous Time

In this part, we present some continuous-time processes and study their self-similar properties and long range dependence. We consider three cases: Gaussian H-sssi models, non-Gaussian H-sssi models and multi-fractal processes.

Gaussian H-sssi Models

The canonical Gaussian H-sssi model is the fractional Brownian motion, which is unique under a scaling constant. We recall first of all the definition of Brownian motion.

Definition 2.2.8. *Let $B(t)$ be a stochastic process with continuous sample paths and such that, $\forall t \in \mathbb{R}$,*

(i) $B(0) = 0\ a.s.$;
(ii) $B(t)\ has\ independent\ increments$;
(iii) $for\ each\ t,\ B(t)\ has\ a\ Gaussian\ distribution\ with\ mean\ zero\ and\ variance\ t$;
(iv) $E[B(t) - B(s)] = 0$;

Then $B(t)$ is called the (standard) Brownian motion.

According to the definition, we can remark that the Brownian motion is a $\frac{1}{2}$-ss process. However, since the increments of the Brownian motion are independent, it is not a long memory process.

Among continuous-time processes, another important model is the fractional Brownian motion, which is the canonical example of Gaussian H-sssi processes.

Definition 2.2.9. *Let $B(t)$ be the standard Brownian motion with $0 < H < 1$. Define $B_H(t)$ the stochastic integral*

$$B_H(t) = \int \omega_H(t, u) dB(u)$$

where the convergence of the integral is under L^2-norm with respect to the Lebesgue measure on the real numbers, where the weight function ω_H satisfying the following conditions:

- $\omega_H(t,u) = 0$, *for* $t \leq u$;
- $\omega_H(t,u) = (t-u)^{H-\frac{1}{2}}$, *for* $0 \leq u < t$;
- $\omega_H(t,u) = (t-u)^{H-\frac{1}{2}} - (-u)^{H-\frac{1}{2}}$, *for* $u < 0$,

then $B_H(t)$ *is called the fractional Brownian motion (fBm) with self-similar parameter* H.

Now we recall some interesting properties concerning the fractional Brownian motion:

1. The fBm is the unique Gaussian H-sssi process, with respect to a scaling constant.
2. The covariance function $\gamma(k)$ $(k \in \mathbb{Z})$ of the fBm is proportional to $|k|^{2H-2}$, as $k \to \infty$.
3. When $0 < H < 1$, the fBm exhibits long memory behavior, since its spectral density $f(w)(-\pi < \omega < \pi)$ is proportional to ω^{-2H-1}, as $\omega \to 0$.
4. We can divide the class of the fBm into three classes: anti-persistent process ($0 < H < \frac{1}{2}$), random walk ($H = \frac{1}{2}$) and persistent process ($\frac{1}{2} < H < 1$).
5. To check whether a process $\{X(t)\}_{t\in\mathbb{R}}$ is the fBm, we need to verify the following items:

   (i) It is a Gaussian process with mean 0, $X(0) = 0$;

   (ii) $E[X^2(t)] = \sigma^2|t|^{2H}$, for some $\sigma > 0$, $0 < H < 1$;

   (iii) The process $\{X(t)\}_{t\in\mathbb{R}}$ has stationary increments.
6. We often consider the fBm as a generalization of the Brownian motion. When $H = 1$, we get the degenerate case.

Since the fBm is the H-sssi processes, it builds a bridge between the continuous-time processes and the discrete-time processes.

Non-Gaussian H-sssi Models

In contrast to the uniqueness of the Gaussian H-sssi process, we have an infinite number of non-Gaussian H-sssi models, among which there is a large class of models with infinite variance – the α-stable processes (with $0 < \alpha < 2$). If $\alpha = 2$, they reduce to the Gaussian processes.

For α-stable H-sssi processes with $0 < \alpha < 2$, we have the following properties:

1. For different $\alpha \in (0,2)$, the parameter H lies in different intervals:
   - $H \in (0, \frac{1}{\alpha}]$, if $0 < \alpha < 1$
   - $H \in (0,1]$, if $1 < \alpha < 2$.

2. When $0 < \alpha < 2$ and $H \neq \frac{1}{\alpha}$, one of the most commonly studied α-stable H-sssi processes is the linear fractal stable motion (or linear fractal Lévy motion).
   When $0 < \alpha < 2$ and $H = \frac{1}{\alpha}$, we get the α-stable Lévy motion.

3. When $0 < \alpha < 1$, the unique non-degenerate α-stable $\frac{1}{\alpha}$-sssi process is the α-stable Lévy motion.
   When $1 \leq \alpha < 2$, there is no uniqueness any more. For example, when $\alpha = 1$, X is a 1-stable random variable. This means that the process $Y(t) = t \cdot X$ $(\forall t \in \mathbb{R})$ is a 1-sssi process.

4. When $1 < \alpha < 2$, the log-fractional stable motion is $\frac{1}{\alpha}$-sssi process.

5. There exist many other standard families of α-stable H-sssi processes, and for simplicity, we concentrate on the symmetric case. The three famous symmetric α-stable ($S\alpha S$) H-sssi processes are:

   - the linear fractional stable motions;
   - the real harmonizable fractional stable motions;
   - the sub-Gaussian fractional motions.

   All these three models reduce themselves to fractional Brownian motion if $\alpha = 2$. The corresponding increments processes are called linear fractional stable noise, real harmonizable fractional stable noise and sub-Gaussian fractional noise.

6. Comparing the fBm and the α-stable Lévy processes, we find that the self-similarity property can have quite different origins:

   - It can arise from strong dependence between increments in the absence of high variability (ex. fBm). This mechanism for self-similarity is called "Joseph effect".
   - It can arise from high variability, whose increments are independent and heavy tailed (α-stable Lévy process). This mechanism for self-similarity is called "Noach effect".

7. The relationship between self-similar processes, Gaussian processes and Lévy processes can be characterized as follows:

   - A self-similar and Gaussian process is the fractional Brownian motion;
   - A self-similar and Lévy process is the so-called α-stable process;
   - A Gaussian and Lévy process corresponds to the Brownian motion with drift;
   - The Brownian motion acts as a particular case of all these three kinds of processes.

Multi-fractal Processes

The notion of self-similarity coincides with the original definition of fractality for geometric objets. The concept of fractality comes from the fractal geometry to describe the self-similar processes. So we can generalize the notion of self-similarity to multi-fractality.

Definition 2.2.10. *A random process $\{X(t)\}_{t>0}$ is called a multi-fractal process if for any $a > 0$, there exits a random function $M(a)$, such that $X(at) =^d M(a)X(t)$.*

Exact self-similar process is a degenerate example of multi-fractal process, which can be called mono-fractal process.

2.2.5 Examples of Self-similar Processes in Discrete Time

In this part, we investigate the classical discrete-time processes and their self-similarity properties. The study concerns fGn, k-factor GARMA process, k-factor GIGARCH process, process with switches, process with breaks and process with threshold.

- Fractional Gaussian Noise (fGn)

  Definition 2.2.11. *A process $(X_t)_{t\in\mathbb{Z}}$ is called a fractional Gaussian noise, or fGn, if it satisfies, for all $t \in \mathbb{Z}$, $X_t = B_H(t) - B_H(t-1)$, where $\{B_H(t)\}_{t\in\mathbb{R}}$ is a fractional Brownian motion introduced in Definition 2.2.9.*

  Thus, the fractional Gaussian noise (fGn) is defined as the increments sequences of the process fBm. And it has the following properties:

  1. The fractional Gaussian noise is an exactly self-similar stationary Gaussian process with zero mean.
  2. Since the fractional Brownian motion is the unique H-sssi Gaussian process, the process fGn is also the unique stationary Gaussian process which is exactly self-similar. The uniqueness is under the sense of a scaling constant.

- k-factor GARMA Processes

  Definition 2.2.12. *A process $(X_t)_t$ is called a k-factor Gegenbauer autoregressive moving average process, or a k-factor GARMA process, if it has the following representation*

  $$\Phi(B)\prod_{i=1}^{k}(1-2\nu_i B+B^2)^{d_i}X_t = \Theta(B)\varepsilon_t, \tag{2.14}$$

  where k is a finite integer, $|\nu_i| \leq 1$ for all $i = 1, \cdots, k$, $(\varepsilon_t)_t$ is a white noise with zero mean and variance σ_ε^2, and $\Phi(B) = I - \phi_1 B - \cdots - \phi_p B^p$, $\Theta(B) = I - \theta_1 B - \cdots - \theta_q B^q$, $d_i \in \mathbb{R}$, B is the backshift operator satisfying $BX_t = X_{t-1}$.

Notice that the frequencies $\lambda_i = \cos^{-1}(\nu_i)(i = 1, \cdots, k)$ are called Gegenbauer frequencies. This model was first studied by Woodward et al. (1998) and Giraitis and Leipus (1995).

In the following, we recall the conditions for a covariance stationary k-factor GARMA process to be long memory in order to apply the result of the lemmas that we obtained.

Lemma 2.2.13. *A k-factor GARMA process is covariance stationary and exhibits long memory behavior when ν_i are distinct, all the roots of the polynomials $\Phi(B)$ and $\Theta(B)$ are distinct and outside the unit circle, and if (i) $0 < d_i < \frac{1}{2}$ and $|\nu_i| < 1$ or if (ii) $0 < d_i < \frac{1}{4}$ and $|\nu_i| = 1$, for $i = 1, \cdots, k$.*

Proposition 2.2.14. *Let $(X_t)_t$ be a k-factor GARMA process, then under the conditions of Lemma 2.2.13, it is asymptotically second-order self-similar. Furthermore, under Gaussianity, it is asymptotically self-similar.*

Proof. A k-factor GARMA process is covariance stationary and long memory under the conditions of Lemma 2.2.13. Thus according to Lemma 2.2.6, a k-factor GARMA process is asymptotically second-order self-similar. Under Gaussianity, it is asymptotically self-similar. □

We note that there are many particular cases of k-factor GARMA process described as follows:

1. When $k = 1$ in Equation (2.14), we get Gegenbauer ARMA (GARMA) process introduced by Gray et al. (1989). This model contains the Gegenbauer process (GI(d)), without short memory terms.
2. When $k = 1$ and $\nu = 1$ in Equation (2.14), we get the fractional integrated ARMA (ARFIMA) process introduced by Granger and Joyeux (1980) and Hosking (1981). This model contains the fractional integrated process (FI(d))when there are no short memory terms.

As a consequence, these models, as the particular cases of k-factor GARMA process, are covariance stationary, they are asymptotically self-similar, for which the results are given in Proposition 2.2.14.

- Heteroscedastic Processes
  GARCH Processes and Related Processes

Definition 2.2.15. *A process $(X_t)_t$ is a generalized autoregressive conditional heteroscedastic process with order p (≥ 0) and q (> 0), or GARCH(p,q), if it has the following representation $X_t = \sigma_t \varepsilon_t$ and*

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i X_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sum_{t-j}^{2} = \alpha_0 + a(B)X_t^2 + b(B)\sigma_t^2, \qquad (2.15)$$

where $\alpha_0 > 0$, $\alpha_i \geq 0 (i = 1, \cdots, q)$, $\beta_j \geq 0 (j = 1, \cdots, p)$, $\{\varepsilon_t\} \sim IID(0,1)$, *and* ε_t *is independent of* $\{X_{t-k}, k \geq 1\}$ *for all* t, B *is the backshift operator,* $a(B)$ *and* $b(B)$ *are polynomials in* B *of order* q *and* p *respectively.*

This model was introduced by Bollerslev (1986). If $\beta_j = 0$ in (2.15), we get an ARCH process (Engle 1982). If $\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j = 1$, we get an IGARCH process (Bollerslev 1988).

Lemma 2.2.16. *The GARCH (p,q) process defined in Definition* 2.2.15 *is covariance stationary with* $E(X_t) = 0$, $Var(X_t) = \alpha_0(1-a(1)-b(1))^{-1}$ *and* $Cov(X_t, X_s) = 0$ *for* $t \neq s$, *if and only if* $a(1) + b(1) < 1$.

See Bollerslev (1986) for more details.

Proposition 2.2.17. *Let* $(X_t)_t$ *be the GARCH process introduced in Definition* 2.2.15, *if* $a(1) + b(1) < 1$, *then it is not asymptotically second-order self-similar.*

Proof. If $a(1) + b(1) < 1$, according to the Lemma 2.2.16, the GARCH model is covariance stationary and short memory. Thus, following the Lemma 2.2.5, it is not asymptotically second-order self-similar. □

The conclusion also works for particular GARCH models like ARCH and IGARCH models.

k-factor GIGARCH Processes

Definition 2.2.18. *A process* $(X_t)_t$ *is called a* k*-factor Gegenbauer integrated generalized autoregressive conditional heteroscedastic process, or a* k*-factor GIGARCH process, if it has the following representation*

$$\Phi(B) \prod_{i=1}^{k} (I - 2\nu_i B + B^2)^{d_i} (X_t - \mu) = \Theta(B)\varepsilon_t, \tag{2.16}$$

where $\varepsilon_t = \xi_t \sigma_t$ *with* $(\xi_t)_{t \in \mathbb{Z}}$ *a white noise process with unit variance and mean zero,* $\sigma_t^2 = a_0 + \sum_{i=1}^{r} a_i \varepsilon_{t-i}^2 + \sum_{j=1}^{s} b_j \sigma_{t-j}$, μ *the mean of the process* $(X_t)_t$, $\Phi(B)$ *and* $\Theta(B)$ *polynomials in* B *of order p and q respectively, B the backshift operator satisfying* $BX_t = X_{t-1}$, $d = (d_1, ..., d_k)$ *the memory parameters and* $\nu = (\nu_1, ..., \nu_k)$ *the frequency location parameters,* $d_i \in \mathbb{R}$, $|\nu_i| \leq 1$, $i = 1, \cdots, k$.

Notice that the frequencies $\lambda_i = cos^{-1}(\nu_i)$ $(i = 1, ..., k)$ are called Gegenbauer frequencies.

Lemma 2.2.19. *A k-factor GIGARCH process $(X_t)_{t\in\mathbb{Z}}$ is covariance stationary when the following hypotheses are satisfied:*

$(\mathbf{H_0})$: $a_0 > 0, a_1, ..., a_r, b_1, ...b_s \geq 0$ *and* $\sum_{i=1}^{r} a_i + \sum_{i=1}^{s} b_i < 1$*;*

$(\mathbf{H_1})$: $|d_i| < 1$ *if* $|\nu_i| < 1$ *and* $|d_i| < \frac{1}{4}$ *if* $\nu_i = 1$*, for* $i = 1, ..., k$*;*

$(\mathbf{H_2})$: *all the roots of* $\Phi(B) = 0$ *and* $\Theta(B) = 0$ *lie outside the unit circle and there is no common root;*

$(\mathbf{H_3})$: $E(\varepsilon_t^4) < \infty$*;*

$(\mathbf{H_4})$: $\nu_i (i = 1, ..., k)$ *are supposed to be known.*

Lemma 2.2.20. *A covariance stationary k-factor GIGARCH process is long memory when $|\nu_i| < 1$ and $0 < d_i < \frac{1}{2}$. The asymptotic behaviors of spectral density and autocorrelation function are as follows: $f_X(\lambda) \sim c(\lambda)|\lambda - \lambda_j|^{-2d}$, as $\lambda \to \lambda_j$, $j = 1, ..., k$, and $\rho_k \sim k^{2d-1}\cos(k\lambda_j)$, as $k \to \infty$, where $\lambda_j = \cos^{-1}(\nu_j)$ is the Gegenbauer frequency.*

See Guégan (2003) for detail.

Proposition 2.2.21. *Let $(X_t)_t$ be a k-factor GIGARCH process, it is asymptotically second-order self-similar under the conditions of Lemma* (2.2.19) *and Lemma* (2.2.20)*. Furthermore, under Gaussianity, it is asymptotically self-similar.*

Proof. A k-factor GIGARCH process is covariance stationary and exhibits long memory under the conditions of Lemma 2.2.19 and Lemma 2.2.20, thus according to Lemma 2.2.6, a k-factor GIGARCH process is asymptotically second-order self-similar. Under the framework of Gaussianity, it is asymptotically self-similar. □

The k-factor GIGARCH processes exhibit both heteroscedasticity and long memory characteristics.

- Processes with Switches and Jumps
  Structural breaks have been observed in many economic and financial time series. A lot of models have been proposed in order to capture the existence of structural changes and complex dynamic patterns.

  Let $(X_t)_t$ be a process whose recursive scheme is

$$X_t = \mu_{s_t} + \varepsilon_t, \tag{2.17}$$

  where $(\mu_{s_t})_t$ is a process that we specify below and $(\varepsilon_t)_t$ is a strong white noise, independent of $(\mu_{s_t})_t$.

  With respect to the process $(\mu_{s_t})_t$, we distinguish two cases:

1. $(\mu_{s_t})_t$ depends on a hidden ergodic Markov chain $(s_t)_t$;
2. If $(\mu_{s_t})_t = \mu_t$, then we will assume that this process depends on a probability p.

The first class of models includes "models with switches" and the second class includes "models with breaks". If these processes are covariance stationary, for most of them, it has been proved that they are short memory processes. Nevertheless, among these processes, some exhibit empirically a kind of long memory behavior when we observe the corresponding behavior of the sample autocorrelation function. Thus for these processes, studying their self-similarity is more complex than for the previous class of processes that we have already discussed.

Processes with Switches and Breaks

We consider a two-state Markov Switching model $(X_t)_t$ defined by the following equations:

$$X_t = \mu_{s_t} + \phi_{s_t} X_{t-1} + \sigma_{s_t} \varepsilon_t, \tag{2.18}$$

where μ_{s_t}, ϕ_{s_t} and σ_{s_t} $(i = 1, 2)$ are real parameters, σ_i are positive and $(\varepsilon_t)_t$ is a strong white noise with mean $m \in \mathbb{R}$ and variance $\sigma^2 \in \mathbb{R}_+^*$, and $(\varepsilon_t)_t$ is independent of the hidden ergodic Markov chain $(s_t)_t$, which is characterized by its transition probabilities p_{ij}, defined by:

$$P[s_t = j | s_{t-1} = i] = p_{ij} \tag{2.19}$$

with $0 \leq p_{ij} \leq 1$ and $\sum_{j=1}^{2} p_{ij} = 1$ $(i = 1, 2)$. Thus the transition matrix is equal to

$$P = \begin{pmatrix} p_{11} & 1 - p_{22} \\ 1 - p_{11} & p_{22} \end{pmatrix}.$$

Proposition 2.2.22. *Let $(X_t)_t$ be a process defined by Equation* (2.18) *and Equation* (2.19)*, then if $\max_{i=1,2}\{p_{i1}|\phi_1|^2 + p_{i2}|\phi_2|^2\} < 1$, the two-state Markov Switching model is covariance stationary and also short memory, thus it is not asymptotically second-order self-similar.*

Proof. Under the above conditions, the process $(X_t)_t$ defined by Equation (2.18) and Equation (2.19) is covariance stationary (Yang 2000). Under stationarity, theoretically this model is known to be short memory. Thus, following Lemma 2.2.5, it is not asymptotically second-order self-similar. □

Actually, there are many other interesting models with switches contained in Equation (2.17), for example:

- the mean regime switching model $X_t = \mu_{s_t} + \varepsilon_t$;
- the mean variance regime switching model $X_t = \mu_{s_t} + \sigma_{s_t}\varepsilon_t$;

– the sign model $X_t = sign(X_{t-1}) + \varepsilon_t,\ where\ \varepsilon_t \sim N(0, \sigma^2)$.

For all of these models with switches, if they are covariance stationary, then they are theoretically short memory processes, although empirically they exhibit long memory behavior. Likewise, they are not asymptotically second-order self-similar.

Now we consider processes with breaks. Assume the process $(X_t)_t$ is defined by

$$X_t = \mu_t + \varepsilon_t, \tag{2.20}$$

where the process $(\mu_{s_t})_t = \mu_t$ depends on a probability p. Different dynamics of the process $(\mu_t)_t$ corresponds to different break models, for example: the Binomial model, the random walk model with a Bernouilli process, the STOPBREAK model, the stationary random level shift model, the mean-plus-noise model, etc. Under stationarity, these models are short memory and is not asymptotically second-order self-similar. However, empirically, the sample autocorrelation function of the Markov-Switching process decreases in an hyperbolic way towards zero, thus it exhibits the long memory behavior. According to Lemma 2.2.6, empirically this process appears asymptotically second-order self-similar. We will call this kind of behavior a "Spurious asymptotically second-order self-similar" behavior. Similarly, when the models with breaks and switches introduced above are covariance stationary, they are also theoretically short memory processes, although empirically they exhibit long memory behavior. Therefore, under stationarity, they are not asymptotically second-order self-similar, but present some "spurious asymptotical second-order self-similar" behavior.

Processes with Threshold

Consider the general form of the processes with threshold:

$$X_t = f(X_{t-1})(1 - G(X_{t-d}, \gamma, c)) + g(X_{t-1})G(X_{t-d}, \gamma, c) + \varepsilon_t, \tag{2.21}$$

where the function f and g can be any linear or nonlinear functions of the past values of X_t or ε_t. The process $(\varepsilon_t)_t$ is a strong white noise and G is an indicator function or a continuous function. For a given threshold c and the position of the random variable X_{t-d} with respect to this threshold c, the process $(X_t)_t$ contains different models, for example, SETAR models, STAR models, etc.

Lemma 2.2.23. *Consider the threshold process defined in Equation* (2.21)*, if functions f and g are to short memory processes, then under covariance stationarity, the model* (2.21) *is short memory.*

Proposition 2.2.24. *For the process with threshold defined in Equation* (2.21)*, if the functions f and g correspond to short memory process, then the model* (2.21) *is not asymptotically second-order self-similar.*

Proof. According to Lemma 2.2.23, if the functions f and g are short memory process, the process (2.21) is short memory. Thus, following Lemma 2.2.5, the process (2.21) is not asymptotically second-order self-similar. □

Now we consider the long memory SETAR model, defined as follows:

$$X_t = (1-B)^{-d}\varepsilon_t^{(1)} I_t(X_{t-d} \leq c) + \varepsilon_t^{(2)}[1 - I_t(X_{t-d} \leq c)], \tag{2.22}$$

and the assumptions:
$(\mathbf{H}_5)$: The process $(\varepsilon_t^{(i)})_t$ $(i = 1, 2)$ is a sequence of independent identically distributed random variables.
$(\mathbf{H}_6)$: The long memory parameter d is such that $0 < d < 1/2$. So in regime 1, the process is invertible and stationary.

Lemma 2.2.25. *Under the assumptions $(\boldsymbol{H}_5)$ and $(\boldsymbol{H}_6)$, the process $(X_t)_t$ defined in Equation* (2.22) *is globally covariance stationary. Asymptotically its spectral density function is such that $f_X(\omega) \sim C\omega^{-2d}$, as $\omega \to 0$, where C is a positive constant, and its autocorrelation function verifies $\gamma_X(h) \sim \dfrac{\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)} h^{2d-1}$, as $h \to 0$.*

Proposition 2.2.26. *Under the assumptions $(\boldsymbol{H}_5)$ and $(\boldsymbol{H}_6)$, the covariance stationary SETAR model defined in Equation* (2.22) *is asymptotically second-order self-similar. Furthermore, under Gaussianity, it is asymptotically self-similar.*

Proof. Under the assumptions $(\mathbf{H}_5)$ and $(\mathbf{H}_6)$, following Lemma 2.2.25, the covariance stationary model defined in Equation (2.22) is long memory process. According to Lemma 2.2.6, it is asymptotically second-order self-similar. Furthermore, under Gaussianity, it is asymptotically self-similar. □

Dufrénot et al. (2005a, 2005b, 2008) have make the applications of the long memory SETAR models on some real data.

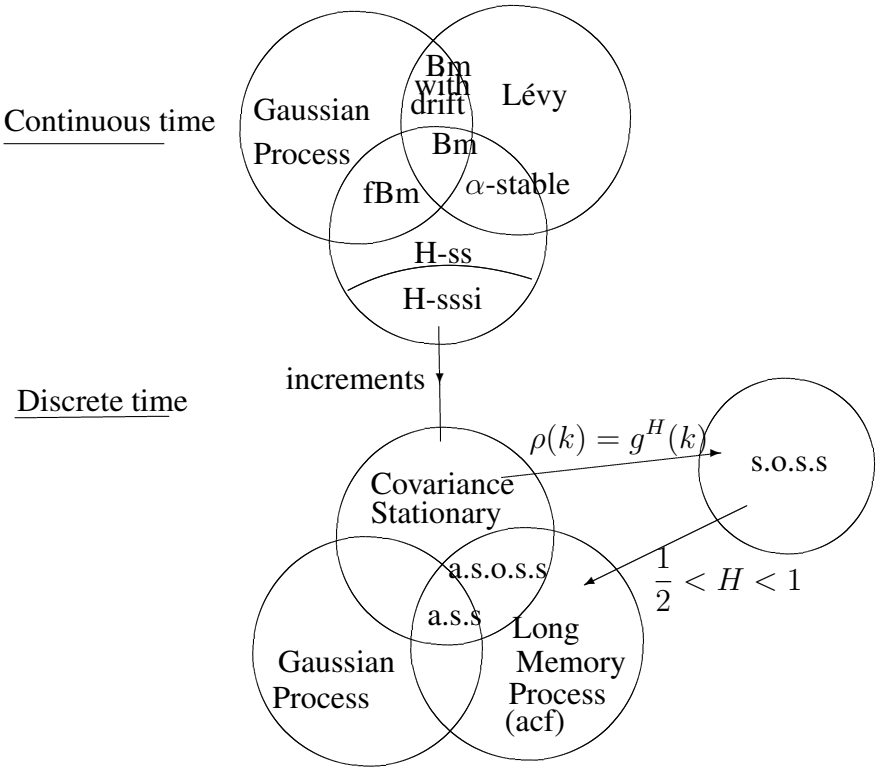
2.2.6 Summarize for the Self-similar Processes

In the previous part, we have discussed the self-similar properties for continuous-time and discrete-time processes. In the continuous-time framework, the fractional Gaussian noise is the unique Gaussian H-sssi process. We made a review for the non-Gaussian H-sssi models. Then we study the multifractal process as one of the generalized self-similar processes. In the discrete-time framework, we proposed two important lemmas which reveal the relationship between the self-similar properties and long/short memory behaviors. We proved that the covariance stationary long memory process is asymptotically second-order self-similar, while the covariance stationary short memory process is not asymptotically second-order self-similar. Then we apply these two lemmas to several stochastic processes: the ARFIMA models, the k-factor GARMA models, GARCH models, k-factor GIGARCH models, processes with switches, processes with breaks and processes with threshold, etc.

Now, Guégan (2007) has proved that, if a process is not globally stationary, but only locally stationary, which means that for instance, for two subsets of the process, the means are different (as in model (2.18)), then the sample autocorrelation function decreases hyperbolically towards zero. According to the definition of long memory process in time

domain, this process appears globally long memory. The arising question is the existence of self-similarity for this process. It seems that even if it is not globally covariance stationary, it can be globally asymptotically second-order self-similar. Thus, we suggest in that case that the process have a spurious asymptotically second-order self-similar behavior.

We summarize the previous discussions in a graph to give some intuitive overall ideas of the main concepts discussed in the previous sections:



In this graph, H-ss stands for H-self-similar process; H-sssi stands for H-self-similar process with stationary increments; a.s.s for asymptotically self-similar; a.s.o.s.s for asymptotically second-order self-similar; s.o.s.s for second-order self-similar; Autocorelation function (ACF) for long memory process defined in sense of autocorrelation function; Bm for Brownian motion; fBm for fractional Brownian motion; α-stable for α-stable process; Lévy for Lévy process.

Chapter 3

Wavelet Techniques for Time Series Analysis

The discrete wavelet transform (DWT) has several appealing qualities that make it a useful tool for time series analysis. Although developed originally with geophysical applications in mind (Goupillaud et al. 1984), the DWT and its variants have found a home in a wide variety of disciplines – geology, atmospheric science, turbulence and applied mathematics, etc. The engineering community quickly realized that common techniques in signal processing were closely related and developed the framework of multiresolution analysis (Mallat, 1989). There are numerous references over the past decades with respect to wavelet analysis. Introductory texts include Ogden (1997), Vidakovic (1999), and Percival and Walden (2000) from a statistical perspective; Vetterli and Kovačević (1995), Burrus et al. (1998), and Mallat (1998) from an engineering perspective; and Strang and Nguyen (1996) and Chui (1992, 1997) from a more mathematical perspective.

This chapter provides a brief discussion of time-frequency representations, including the classical Fourier transform and short-time Fourier transform. And a detailed introduction of the wavelet transform is followed. This part of review serves for the work in Chapter 4 and Chapter 5 and to make this thesis self-contained.

3.1 Introduction of the Time-frequency Representations

3.1.1 Fourier Transform

The discrete Fourier transform (DFT) may be derived from a variety of perspectives (e.g. approximating the Fourier transform of a function, approximating the Fourier coefficients of a function) (Briggs and Henson, 1995). We prefer to take the viewpoint of approximating a discretely sampled time series $(X_t)_t$ via a linear combination of sines and cosines. Each of these sines and cosines is itself a function of frequency, and therefore the DFT may be seen as a decomposition on a frequency-by-frequency basis. The Fourier basis functions (sines and cosines) are very appealing when representing a stationary time series.

The Fourier transform is an alternative representation of the original time series such that it summarizes information in the data as a function of frequency and therefore does not preserve information in time. This is the opposite of how we observed the original time series, where no frequency resolution was provided. When observing the time series in the time domain, we have complete time resolution and no frequency resolution, whereas the opposite is true after performing the Fourier transform.

3.1.2 Short-Time Fourier Transform

Gabor (1946) recognized that the Fourier transform goes too far by eliminating all time resolution by frequency resolution and attempted to achieve a balance between time and frequency by sliding a window across the time series and taking the Fourier transform of the windowed series. This is known as the Gabor transform or Short-Time Fourier Transform (STFT). The resulting expansion is a function of two parameters, frequency and time-shift. The key property is that the window size is fixed with respect to frequency. This produces a rectangular partitioning of the time-frequency plane.

The choice of window is very important with respect to the performance of the STFT in practice. Since the STFT is simply applying the Fourier transform to pieces of the time series of interests, it will not be able to resolve the events if they happen to appear within the width of the window. In this case, the lack of time resolution of the Fourier transform is present. In general, one cannot achieve simultaneous time and frequency resolution because of the Heisenberg uncertainty principle. In the field of particle physics, an elementary particle does not simultaneously have a precise position and momentum. The better one determines the position of the particle, the less precisely the momentum is known at that time, and vice versa. For signal processing, this rule translates into the fact that signal does not simultaneously have a precise location in time and precise frequency. An interesting discussion of the Heisenberg uncertainty principle in signal processing can be found in Hubbard (1996) and Mallat (1989).

3.1.3 Wavelet Transform

What is needed to overcome the fixed time-frequency partition is a new set of basis functions. The wavelet transform utilizes a basic function (called the mother wavelet), then dilates and translates it to capture the features that are local in time and local in frequency. The resulting time-frequency partition corresponding to the wavelet transform is long in time when capturing low-frequency events, thus having good frequency resolution for these events, and long in frequency when capturing high-frequency events, thus having good time resolution for these events. Consequently, the wavelet transform intelligently adapts itself to capture features across a wide range of time and frequencies.

The Fourier transform, and also the STFT, is a function of frequency, while the wavelet transform is a function of scale. Actually, the scale in a wavelet transform is indeed related to frequency. Loosely speaking, scale is inversely proportional to a frequency interval. If the scale parameter increases, then the wavelet basis is

- stretched in the time domain,
- shrunk in the frequency domain, and
- shifted toward lower frequencies.

Conversely, a decrease in the scale parameter

- reduces the time support,
- increases the number of frequencies captured, and
- shifts toward higher frequencies.

3.2 Properties of the Wavelet Transform

Although the Fourier transform, the STFT, and wavelet transform are all alternative representations, each transform is well suited to certain types of applications. Wavelets, with their ability to quantify events in both time and scale, are well suited to study a wide range of time series. We introduce some features of wavelets and wavelet transforms before discussing the discrete wavelet transform.

3.2.1 Continuous Wavelet Functions

A wavelet $\psi(t)$ is simply a function of time t that obeys a basic rule, known as the wavelet admissibility condition:

$$C_\psi = \int_0^\infty \frac{|\Psi(f)|}{f} df < \infty, \tag{3.1}$$

where $\Psi(f)$ is the Fourier transform, a function of frequency f, of $\psi(t)$. This condition ensures that $\Psi(f)$ goes to zero quickly as $f \to 0$ (Grossmann and Morlet, 1984; Mallat, 1998). In fact, to guarantee that $C_\psi < \infty$ we must impose $\Psi(0) = 0$, which is equivalent to

$$\int_{-\infty}^{\infty} \psi(t) dt = 0. \tag{3.2}$$

A second condition imposed on a wavelet function is unit energy, that is

$$\int_{-\infty}^{\infty} |\psi(t)|^2 dt = 1. \tag{3.3}$$

By satisfying both Equation (3.2) and (3.3), the wavelet function must have nonzero entries, but all departures from zero must cancel out. The classic example of a continuous-time wavelet function is the Morlet wavelet. The Mexican hat wavelet is an example of symmetric wavelet function, see Percival and Walden (2000) for more details.

The continuous wavelet transform (CWT) is a function of two variables $W(u, s)$ and

is obtained by simply projecting the function of interest $x(t)$ onto a particular wavelet ψ via

$$W(u,s) = \int_{-\infty}^{\infty} x(t)\psi_{u,s}(t)dt, \tag{3.4}$$

where

$$\psi_{u,s}(t) = \frac{1}{\sqrt{s}}\psi(\frac{t-u}{s})$$

is the translated (by u) and dilated (by s) version of the original wavelet function. The resulting wavelet coefficients are a function of these two parameters, location and scale, even though the original function is only a function of one parameter. By applying shifted and translated versions of the mother wavelet to a function of interest, we are breaking down the complicated structure present in the function into simpler components. This is called analyzing and decomposing the function. If a wavelet satisfies the admissibility condition (Equation (3.1)), then an inverse operation may be performed to produce the function from its wavelet coefficients; that is,

$$x(t) = \frac{1}{C_\psi}\int_0^{\infty}\int_{-\infty}^{\infty} W(u,s)\psi_{u,s}(t)du\frac{ds}{s^2}.$$

This is called synthesizing and reconstructing the function. A key property of wavelet transforms is their ability to decompose and perfectly reconstruct a square-integrable function.

3.2.2 Continuous versus Discrete Wavelet Transform

As mentioned previously, the CWT is a function of two parameters and therefore contains a high amount of extra (redundant) information when analyzing a function. Enough information would be available to easily detect this discontinuity if one were to sample only a portion of the CWT. Thus we reduce our task from analyzing an image (the CWT) with continuous parameters u and s to viewing a small number of scales with a varying number of wavelet coefficients at each scale. This is the discrete wavelet transform (DWT). Although the DWT may be derived without referring to the CWT, we may review it as a "discretization" of the CWT through sampling specific wavelet coefficients (Vidakovic, 1999). A critical sampling of the CWT is obtained via

$$s = 2^{-j}\ and\ u = k2^{-j},$$

where j and k are integers representing the set of discrete translations and discrete dilations. A critical sampling defines the resolution of the DWT in both time and frequency. We use the term critical sampling to denote the minimum number of coefficients sampled from the CWT to ensure that all information present in the original function is retained by the wavelet coefficients. The CWT takes on values at every point on the time-frequency plane. The DWT, on the other hand, only takes on values at very few points and the wavelets that follow these values $\psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k)$ (for all integers j, k), produce an orthogonal basis. If we select a sequence of dyadic scales and instead use all integer translations:

$$s = 2^{-j}\ and\ u = k,$$

then we arrive at the maximal overlap DWT (MODWT).

Over the past decades, both the DWT and MODWT have been utilized in a variety of fields and under a variety of names. With respect to economics and finance, J.B. Ramsey and co-authors first introduced wavelets into the mainstream literature. Ramsey and Zhang (1997) performed a time-frequency analysis of foreign exchange rates (Deutsche Mark versus U.S. Dollar and Japanese yen versus U.S. Dollar) using wavelets. They found that wavelet analysis succinctly captured a variety of non-stationary events in the series. Ramsey and Lampart (1998a, b) decomposed economic variables across several wavelet scales in order to identify different relationships between money and income, and between consumption and income. See Ramsey (1999) for a review article on wavelets in economics and finance. Stengos and Sun (2001) designed a consistent specification test for a regression model based on wavelet estimation, and Fan (2000) proposed a wavelet estimator of a partial linear model by regressing boundary independent DWT coefficients of the dependent variable on the corresponding DWT coefficients of the regressors in linear part of the model across several scales.

3.3 Discrete Wavelet Filters

Before formulating the DWT or MODWT, we discuss the discrete wavelet filters available to us. Fundamental properties of the continuous wavelet functions (filters), such as integration to zero and unit energy (Equation (3.2) and (3.3)), have discrete counterparts.

Let $h_l = (h_0, \cdots, h_{L-1})$ be a finite length discrete wavelet filter such that it integrates (sums) to zero

$$\sum_{l=0}^{L-1} h_l = 0 \tag{3.5}$$

and have unit energy

$$\sum_{l=0}^{L-1} h_l^2 = 1. \tag{3.6}$$

In addition to Equation (3.5) and (3.6), the wavelet (or high-pass) filter h_l is orthogonal to its even shifts; that is,

$$\sum_{l=0}^{L-1} h_l h_{l+2n} = 0, \quad for\ all\ nonzero\ integers\ n. \tag{3.7}$$

This comes from the relationship between the DWT and CWT via a critical sampling. To construct the orthonormal matrix that defines the DWT, wavelet coefficients cannot interact with one another. Equations (3.6) and (3.7) may be succinctly expressed in the frequency domain via the squared gain function

$$\mathcal{H}(f) + \mathcal{H}(f + \frac{1}{2}) = 2, \ for\ all\ f. \tag{3.8}$$

The natural object to complement a high-pass filter is a low-pass (scaling) filter whose squared gain function monotonically increases as $f \to 0$. By applying both h_l and g_l to an observed time series, we separate high-frequency oscillations from low-frequency ones. We will denote a low-pass filter as $g_l = (g_0, \cdots, g_{L-1})$, and its transfer and squared gain function are given by $G(f)$ and $\mathcal{G}(f)$, respectively. For all the wavelets considered here, the low-pass filter coefficients are determined by the "quadrature mirror relationship":

$$g_l = (-1)^{l+1} h_{L-1-l}, \quad for\ l = 0, \cdots, L-1. \tag{3.9}$$

The inverse relationship is given by $h_l = (-1)^l g_{L-1-l}$. Using Equation (3.9) allows us to relate the transfer function for g_l and h_l via

$$G(f) = \sum_{l=0}^{L-1} g_l e^{-i2\pi f} = e^{-i2\pi f(L-1)} H(f - \frac{1}{2}),$$

and, thus the squared gain function follows

$$\mathcal{G}(f) = |e^{-2\pi f(L-1)} H(f - \frac{1}{2})|^2 = |H(\frac{1}{2} - f)|^2 = \mathcal{H}(\frac{1}{2} - f). \tag{3.10}$$

Finally, a band-pass filter has a squared gain function that covers an interval of frequencies and then decays to zero as $f \to 0$ and $f \to \frac{1}{2}$. We may construct a band-pass filter by recursively applying a combination of low-pass and high-pass filters.

3.3.1 Haar Wavelets

The first wavelet filter, the Haar wavelet (Haar, 1910), remained in relatively obscurity until the convergence of several disciplines to form what we now know in a broad sense as wavelet methodology. It is a filter of length $L = 2$ that can be succinctly defined by its scaling (low-pass) filter coefficients

$$g_0 = g_1 = \frac{1}{\sqrt{2}},$$

or equivalently by its wavelet (high-pass) filter coefficients $h_0 = \frac{1}{\sqrt{2}}$ and $h_l = -\frac{1}{\sqrt{2}}$ through the inverse quadrature mirror relationship.

The Haar wavelet is special since it is the only symmetric compactly supported orthonormal wavelet (Daubechies, 1992). It is also useful for presenting the basic properties shared by all the wavelet filters. Although the Haar wavelet filter is easy to visualize and implement, it is inadequate for most real-world applications in that it is a poor approximation to an ideal band-pass filter. An ideal band-pass filter is proportional to one inside the desired frequency interval and zero at all other frequencies.

3.3.2 Daubechies Wavelets

The Daubechies wavelet filters represent a collection of wavelets that improve the frequency-domain characteristics of the Haar wavelet and may be still interpreted as generalized differences of adjacent averages (Daubechies, 1992). Daubechies derived these wavelets from the criteria of a compactly supported function with the maximum number of vanishing moments. In general, there are no explicit time-domain formulae for this class for wavelet filters (when possible, filter coefficients will be provided). Daubechies first chose an extremal phase factorization, whose resulting wavelet we denote by $D(L)$ where L is the length of the filter. An alternative factorization leads to the least asymmetric class of wavelets, which we denote by $LA(L)$. Shann and Yen (1999) provided exact values for both the extremal phase and least asymmetric wavelets of length $L \in \{8, 10\}$. Longer extremal phase and least asymmetric wavelet filters do not have a closed form and have been tabulated by, for example, Daubechies (1992) and Percival and Walden (2000).

For Daubechies wavelets, the number of vanishing moments is half the filter length, thus the Haar wavelet has a single vanishing moment, the $D(4)$ wavelet has two vanishing moments, and the $D(8)$ and $LA(8)$ wavelets both have four vanishing moments. One implication of this property is that longer wavelet filters may produce stationary wavelet coefficients vectors from "higher degree" non-stationary stochastic processes.

3.3.3 Minimum Bandwidth Discrete-time Wavelets

The minimum-bandwidth discrete-time (MBDT) wavelets are a new class of orthonormal wavelet filters that were developed by Morris and Peravali (1999). They are generated via an iterative optimization of the spectral factorization procedure. This results in a family of filters that have similar values to those of the Daubechies wavelets, but are obtained through a completely iterative procedure. Hence, although Daubechies wavelets have closed form expressions for the squared gain functions, the MBDT wavelets do not.

What MBDT wavelets offer is an improved approximation to an ideal band-pass filter for length $L \geq 8$. The MBDT wavelets offer superior frequency domain properties to Daubechies wavelets given a filter of the same length. From a statistical point of view, band-pass filtering a time series is important to produce approximately uncorrelated wavelet coefficients for processes with quite general spectra. The MBDT wavelets offer improved frequency resolution to that of the Daubechies family of wavelets. They are most similar to the Daubechies extremal phase family of wavelets.

3.4 Discrete Wavelet Transform (DWT)

These days, data availability is becoming less and less of a problem. For instance, most of the exchanges and especially those that trade electronically would gladly provide tick-by-tick data to interested parties. Data vendors have themselves improved their data structures and provide their users with tools to collect data from over-the-counter (OTC) markets. Data vendors like Reuters, for instance, transmit more than 275,000 prices per day

for foreign exchange spot rates alone. With such massive amounts of financial data being collected at any given time, the discrete wavelet transform is computationally cheaper relative to the continuous wavelet transform. Furthermore, economic and financial data are inherently discrete, and thus we consider discrete transformations.

The DWT is an alternative to the Fourier transform for analyzing a time series. It provides wavelet coefficients that are local in both time and frequency. The DWT may be thought of as a critical sampling of the CWT–that is, it uses the least number of coefficients possible. The DWT possesses key attributes such as approximately decorrelating certain processes and efficiently captures important features of a process in a limited number of coefficients.

There are a variety of ways to express the basic DWT. As we have seen from the definition of the CWT, wavelet coefficients were obtained from projecting (or correlating) portions of a time series with translated and dilated versions of the wavelet function. We also note that projection is closely related to convolution and we introduce filtering concepts for discrete series. Here it is most straightforward to introduce the DWT through a simple matrix operation.

Let $\mathbf{x}$ be a dyadic length vector ($N = 2^J$) of observations. The length N vector of discrete wavelet coefficients $\mathbf{w}$ is obtained via

$$\mathbf{w} = \mathcal{W}\mathbf{x},$$

where $\mathcal{W}$ is an $N \times N$ orthonormal matrix defining the DWT, composed of the wavelet and scaling filter coefficients arranged on a row-by-row basis. The vector of wavelet coefficients may be organized into $J + 1$ vectors,

$$\mathbf{w} = [\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_J, \mathbf{v}_J], \tag{3.11}$$

where $\mathbf{w}_j$ is a length $N/2^j$ ($j = 1, 2, \cdots, J$) vector of wavelet coefficients associated with changes on a scale of length $\lambda_j = 2^{j-1}$ and $\mathbf{w}_J$ is a length $N/2^J$ vector of scaling coefficients associated with averages on a scale of length $2^J = 2\lambda_J$.

3.4.1 Implementation of the DWT: Pyramid Algorithm

In practice, the DWT is implemented via pyramid algorithm (Mallat, 1989) that, starting with the data x_t, filters a series using h_1 and g_1, subsamples both filter outputs to half their original lengths, keeps the subsampled output from the h_1 filter as wavelet coefficients, and then repeats the above filtering operations on the subsampled output from the g_1 filter.

For each iteration of pyramid algorithm, we require three objects: the data vector $\mathbf{x}$, the wavelet filter h_l, and the scaling filter g_l. Assuming those quantities are passed into the program, the first iteration of the pyramid algorithm begins by filtering (convolving) the data with each filter to obtain the following wavelet and scaling coefficients:

$$w_{1,t} = \sum_{l=0}^{L-1} h_l x_{2t+1-l \ mod \ N} \ \ and \ \ v_{1,t} = \sum_{l=0}^{L-1} g_l x_{2t+1-l \ mod \ N},$$

where $t = 0, 1, \cdots, N/2 - 1$ (this may be combined into the same loop for computational efficiency). Note that the downsampling operation has been included in the filtering step through the subscript of x_t. The N length vector of observations has been high-pass and low-pass filtered to obtain $N/2$ coefficients associated with this information. The second step of the pyramid algorithm starts by defining the "data" to be the scaling coefficients $\mathbf{v}_1$ from the first iteration and apply the filtering operations as above to obtain the second level of wavelet and scaling coefficients :

$$w_{2,t} = \sum_{l=0}^{L-1} h_l v_{1,2t+1-l \ mod \ N} \ \ and \ \ v_{2,t} = \sum_{l=0}^{L-1} g_l v_{1,2t+1-l \ mod \ N},$$

$t = 0, 1, \cdots, N/4 - 1$. Keeping all vectors of wavelet coefficients, and the final level of scaling coefficients, we have the following length N decomposition $\mathbf{w} = [\mathbf{w}_1, \mathbf{w}_2, \mathbf{v}_2]^T$. After the third iteration of the pyramid algorithm, where we apply filtering operations to $\mathbf{v}_2$, the decomposition now looks like $\mathbf{w} = [\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3, \mathbf{v}_3]^T$. The procedure may be repeated up to J times where $J = \log_2(N)$ and gives the vector of wavelet coefficients in Equations (3.11).

Inverting the DWT is achieved through upsampling the final level of wavelet and scaling coefficients, convolving them with their respective filters (wavelet for wavelet and scaling for scaling) and adding up to the two filtered vectors. Starting with the final level of the DWT, upsampling the vectors $\mathbf{w}_J$ and $\mathbf{v}_J$ will result in two new vectors:

$$\mathbf{w}_J^0 = [0 \ w_{J,0}]^T \ \ and \ \ \mathbf{v}_J^0 = [0 \ v_{J,0}]^T.$$

The level $J - 1$ vector of scaling coefficients $\mathbf{v}_{J-1}$ is given by

$$v_{J-1,t} = \sum_{l=0}^{L-1} h_l w^0_{J,t+l \ mod \ 2} + \sum_{l=0}^{L-1} g_l v^0_{J,t+l \ mod \ 2},$$

$t = 0, 1$. Notice that the length of $\mathbf{v}_{J-1}$ is twice that of $\mathbf{v}_J$, as to be expected. The next step of reconstruction involves upsampling to produce

$$\mathbf{w}_{J-1}^0 = [0 \ w_{J-1,0} \ 0 \ w_{J-1,1}]^T \ \ and \ \ \mathbf{v}_{J-1}^0 = [0 \ v_{J-1,0} \ 0 \ v_{J-1,1}]^T,$$

and the level $J - 2$ vector of scaling coefficients $\mathbf{v}_{J-2}$ is given by

$$v_{J-2,t} = \sum_{l=0}^{L-1} h_l w^0_{J-1,t+l \ mod \ 4} + \sum_{l=0}^{L-1} g_l v^0_{J-1,t+l \ mod \ 4},$$

$t = 0, 1, 2, 3$. This procedure may be repeated until the first level of wavelet and scaling coefficients have been upsampled and combined to produce the original vector of observations; that is,

$$x_t = \sum_{l=0}^{L-1} h_l w^0_{1,t+l \ mod \ N} + \sum_{l=0}^{L-1} g_l v^0_{1,t+l \ mod \ N},$$

$t = 0, 1, \cdots, N - 1$.

3.4.2 Multiresolution Analysis

Using the DWT, we may formulate an additive decomposition of a series of observations. Let $\mathbf{d}_j = \mathcal{W}_j^T \mathbf{w}_j$ for $j = 1, \cdots, J$, define the j-th level wavelet detail associated with changes in $\mathbf{x}$ at scale λ_j. The wavelet coefficients $\mathbf{w}_j = \mathcal{W}_j \mathbf{x}$ represent the portion of the wavelet synthesis attributable to scale λ_j. For a length $N = 2^J$ vector of observations, the final wavelet detail $\mathbf{d}_{J+1} = v_J^T \mathbf{v}_J$ is equal to the sample mean of the observations.

A multiresolution analysis (MRA) may now be defined via

$$x_t = \sum_{j=1}^{J+1} d_{j,t}, \;\; t = 0, \cdots, N-1. \tag{3.12}$$

That is, for each observation x_t is a linear combination of wavelet detail coefficients. Let $\mathbf{s}_j = \sum_{k=j+1}^{J+1} \mathbf{d}_k$ define the j-th level wavelet smooth for $0 \leq j \leq J$, where $\mathbf{s}_{J+1}$ is defined to be a vector of zeros. Whereas the wavelet detail $\mathbf{d}_j$ is associated with variations at a particular scale, $\mathbf{s}_j$, is a cumulative sum of these variations and will be smoother and smoother as j increases. In fact $\mathbf{x} - \mathbf{s}_j = \sum_{k=1}^{j} \mathbf{d}_j$ so that only lower-scale details (high-frequency features) will be apparent. The jth level wavelet rough characterizes the remaining lower-scale details through $\mathbf{r}_j = \sum_{k=1}^{j} \mathbf{d}_k$ for $1 \leq j \leq J+1$, where $\mathbf{r}_0$ is defined to be a vector of zeros. A vector of observations may thus be decomposed through a wavelet smooth and rough via

$$\mathbf{x} = \mathbf{s}_j + \mathbf{r}_j,$$

for all j. The terminology "detail" and "smooth" were used by Percival and Walden (2000) to describe additive decompositions from Fourier and wavelet transforms.

3.5 Maximal Overlap Discrete Wavelet Transform (MODWT)

An alternative wavelet transform–the maximal overlap discrete wavelet transform (MODWT)–is computed without subsampling the filtered output. The MODWT gives up orthogonality in order to gain features the DWT does not possess. A consequence of this is that the wavelet and scaling coefficients must be rescaled in order to retain the variance preserving property of the DWT.

Although not an orthonormal transform, the MODWT has several advantages over the DWT such as translation invariance, approximation of a zero-phase filtering operation and easy computation for any sample size (Percival and Walden, 2000). The MODWT may be interpreted as applying the rescaled wavelet filters of the DWT to the vector but not decimating (downsampling) the output after filtering.

The following properties are important in distinguishing the MODWT from the DWT (Percival and Mofjeld, 1997):

1. The MODWT can handle any sample size N, while the J_p-th order partial DWT restricts the sample size to a multiple of 2^{J_p}.

2. The detail and smooth coefficients of a MODWT multiresolution analysis are associated with zero phase filters. This means events that feature in the original time series may be properly aligned with features in the multiresolution analysis.

3. The MODWT is invariant to circularly shifting the original time series. Hence, shifting the time series by an integer unit will shift the MODWT wavelet and scaling coefficients the same amount. This property does not hold for the DWT.

4. While both the DWT and MODWT can perform an analysis of variance on a time series, the MODWT wavelet variance estimator is asymptotically more efficient than the same estimator based on the DWT (Percival, 1995).

The MODWT goes by several names in the statistical and engineering literature, such as, the "stationary DWT" (Nason and Silverman, 1995), "translation invariant DWT" (Coifman and Donoho, 1995; Liang and Parks, 1996), and "time-invariant DWT" (Pesquet et al., 1996).

3.5.1 Definition and Implementation of MODWT

Let $\mathbf{x}$ be an arbitrary length N vector of observations. The length $(J+1)N$ vector of MODWT coefficients $\widetilde{\mathbf{w}}$ is a $(J+1)N \times N$ matrix defining the MODWT. The vector of MODWT coefficients may be organized into $J+1$ vectors via

$$\widetilde{\mathbf{w}} = [\widetilde{\mathbf{w}}_1, \widetilde{\mathbf{w}}_2, \cdots, \widetilde{\mathbf{w}}_J, \widetilde{\mathbf{v}}_J]^T, \tag{3.13}$$

where $\widetilde{\mathbf{w}}_j$ is a length $N/2^j$ vector coefficients associated with changes on a scale of length $\lambda_j = 2^{j-1}$ and $\widetilde{\mathbf{v}}_J$ is a length $N/2^J$ vector of scaling coefficients associated with averages on a scale of length $2^J = 2\lambda_J$, just as with the DWT. For time series of dyadic length ($N = 2^J$), the MODWT may be subsampled and rescaled to obtain DWT wavelet coefficients via

$$w_{j,t} = 2^{j/2}\widetilde{w}_{j,2^j(t+1)-1}, \quad t = 0, \cdots, N/2^j - 1,$$

and DWT scaling coefficients via

$$v_{J,t} = 2^{J/2}\widetilde{v}_{J,2^J(t+1)-1}, \quad t = 0, \cdots, N/2^J - 1.$$

3.5.2 Multiresolution Analysis

An analogous MRA to that of the DWT may be performed utilizing the MODWT via

$$x_t = \sum_{j=1}^{J+1} \widetilde{d}_{j,t}, \;\; t = 0, \cdots, N-1,$$

where $\widetilde{d}_{j,t}$ is the t-th element of $\widetilde{\mathbf{d}}_j = \widetilde{\mathcal{W}}_j^T \widetilde{\mathbf{w}}_j$ for $j = 1, \cdots, J$. We may also define respectively the MODWT-based wavelet smooths and roughs to be

$$\widetilde{s}_{J,t} = \sum_{k=j+1}^{J+1} \widetilde{d}_{k,t} \ \ and \ \ \widetilde{r}_{j,t} = \sum_{k=1}^{j} \widetilde{d}_{k,t}, \ t = 0, \cdots, N-1.$$

A key feature of an MRA using the MODWT is that the wavelet details and smooth are associated with zero-phase filters. Thus, interesting features in the wavelet details and smooth may be perfectly aligned with the original time series. This attribute is not available through the DWT since it subsamples the output of its filtering operations.

3.6 Discrete Wavelet Packet Transform (DWPT)

The DWT has a very specific band-pass filtering structure that partitions the spectrum of long memory process finer and finer as $f \to 0$ (i.e. where the spectrum is unbounded) which is described in detail in the next section. This is done through a succession of filtering and downsampling operations. In order to exploit the approximate decorrelation property for long memory processes with seasonality, we need to generalize the partition scheme of the DWT.

This is easily obtained by performing the discrete wavelet packet transform (DWPT) on the process, see for example, Wickerhauser (1994) and Percival and Walden (2000). Instead of one particular filtering sequence, the DWPT executes all possible filtering combinations to obtain a wavelet packet tree, denoted by $\mathcal{T}$. Let $\mathcal{T} = \{(j,n) | j = 0, \cdots, J; n = 0, \cdots, 2^j - 1\}$ be the collection of all doublets (j,n) that form the indices of the nodes of a wavelet packet tree. An orthonormal basis $\mathcal{B} \subset \mathcal{T}$ is obtained when a collection of DWPT coefficients is chosen, whose ideal band-pass frequencies are disjoint and cover $[0, 1/2]$. The set $\mathcal{B}$ is simply a collection of doublets (j,n) that corresponds to an orthonormal basis. Ramsey and Zhang (1996) used a similar, but more extensive, wave-form dictionary to analyze the Standard and Poor's 500 stock index. They found that this decomposition brought out the intermittent nature of the stock market index, that of intense bursts of activity across a wide frequency range followed by periods of relative quiet.

Let $\mathbf{x}$ be a dyadic length vector $(N = 2^J)$ of observations, then the length N vector of DWPT coefficients $\mathbf{w_B}$ is obtained via

$$\mathbf{w}_{\mathcal{B}} = \mathcal{W}_{\mathcal{B}} \mathbf{x},$$

where $\mathcal{W}_{\mathcal{B}}$ is an $N \times N$ orthonormal matrix defined by the orthonormal basis $\mathcal{B}$. All $(J+1)N$ wavelet packet coefficients may be computed by constructing an overcomplete matrix $\mathcal{W}_{\mathcal{T}}$ and applying it to the vector of observations; that is, $\mathbf{w}_{\mathcal{T}} = \mathcal{W}_{\mathcal{T}} \mathbf{x}$, where $\mathcal{W}_{\mathcal{T}}$ is an $(J+1)N \times N$ matrix.

Constructing the matrices $\mathcal{W}_{\mathcal{B}}$ or $\mathcal{W}_{\mathcal{T}}$ involves a minor amount of book-keeping in order to retain the sequency ordering of the wavelet packet filters. Let $h_0, \cdots, h_{L-1}$ be the

unit scale wavelet (high-pass) filter coefficients from a Daubechies compactly supported wavelet family of even length L, with scaling (low-pass) coefficients computed via the quadrature mirror relationship. Now we define

$$u_{n,l} = \begin{cases} g_l, & if\ n\ mod\ 4 = 0\ or\ 3; \\ h_l, & if\ n\ mod\ 4 = 1\ or\ 2, \end{cases}$$

to be the appropriate filter at a given node of the wavelet packet tree. The wavelet packet filters $u_{n,l}$ preserve the ordering of the DWPT by increasing frequency.

Let $h_0, \cdots, h_{L-1}$ be the unit scale wavelet (high-pass) filter. Thus, the scaling (low-pass) coefficients may be computed via the quadrature mirror relationship:

$$g_l = (-1)^{l+1} h_{L-l-1}, \ \ l = 0, 1, \cdots, L-1.$$

Instead of one particular filtering sequence, the DWPT executes all possible filtering combinations to construct a wavelet packet tree, denoted by $\mathcal{T} = \{(j,n) | j = 0, \cdots, J-1; n = 0, \cdots, 2^j - 1\}$.

The DWPT coefficients are then calculated using the pyramid algorithm of filtering and downsampling (Mallat, 1999). Denote $W_{j,\,n,\,K}$ the $K-$th element of length $N_j (= N/2^j)$, corresponding to the wavelet coefficient vector $\mathbf{W}_{j,n}$, $(j,n) \in \mathcal{T}$ with $\mathbf{W}_{0,\,0} = \mathbf{x}$. Given the DWPT coefficients $W_{j-1,\,[\frac{n}{2}],\,K}$, where $[\,\cdot\,]$ represents the "integer part" operator, then the coefficient $W_{j,\,n,\,K}$ is calculated by

$$W_{j,\,n,\,K} \equiv \sum_{l=0}^{L_j - 1} u_{n,\,l} W_{j-1,\,[\frac{n}{2}],\,2K+1-l\ mod\ N_{j-1}}\ , \ \ K = 0, 1, \cdots, N_j - 1, \tag{3.14}$$

where $L_j = (2^j - 1)(L-1) + 1$ is the length of a level j wavelet filter.

An adaptive orthonormal basis $\mathcal{B} \subset \mathcal{T}$ is obtained when a collection of DWPT coefficients is retained such that band-pass frequencies are disjoint and cover the frequency interval $[0, \frac{1}{2}]$ (Percival and Walden, 2000; Gençay et al., 2001).

As with the DWT, the DWPT is most efficiently computed using a pyramid algorithm (Wickerhauser, 1994). The algorithm has $O(N \log N)$ operations, like the fast Fourier transform (FFT).

There is no longer a convenient interpretation of wavelet packet coefficient vectors with differences at various scales. Instead, the vector $\mathbf{w}_{j,n}$ is associated with the frequency interval $\lambda_{j,n} = [\frac{n}{2^{j+1}}, \frac{n+1}{2^{j+1}}]$. The DWPT coefficient vectors corresponding to the DWT coefficients are given by $\mathbf{w}_{j,1}$ at each scale.

3.7 Maximal Overlap Discrete Wavelet Packet Transform (MODWPT)

Definition of the maximal overlap DWPT (MODWPT) is straightforward and follows directly from the DWPT. The rescaled wavelet packet filter is defined to be $\tilde{u}_{n,l} = u_{n,l}/2^{1/2}$. Using $\tilde{u}_{n,l}$ instead of the filter $u_{n,l}$ and not subsampling the filter output produces the MODWPT coefficients. Hence the vector of MODWPT coefficients $\widetilde{\mathbf{w}}_{j,n}$ is computed recursively given $\widetilde{\mathbf{w}}_{j-1,\frac{n}{2}}$ via

$$\widetilde{w}_{j,n,t} = \sum_{l=0}^{L_j-1} \tilde{u}_{n,l}\widetilde{w}_{j-1,[\frac{n}{2}],t-2^{j-1}l \ mod \ N}, \quad t = 0, 1, \cdots, N-1.$$

Each vector of MODWPT coefficients has length N (to begin the recursion simply define $\widetilde{\mathbf{w}}_{0,0} = \mathbf{x}$). This formulation leads to the efficient computation using a pyramid-type algorithm (Percival and Walden, 2000). As with the DWPT, the MODWPT is an energy preserving transform and we may define the decomposition of energy at each level of the transform via

$$\| \mathbf{x} \|^2 = \sum_{n=0}^{2^j-1} \| \widetilde{\mathbf{w}}_{j,n} \|^2, \quad for \ j = 1, \cdots, J.$$

This corresponds to the basis $\mathcal{B}_j = \{(j,n) | n = 0, \cdots, 2^j - 1\}$. Given a particular level j of the transform, we may also reconstruct $\mathbf{x}$ by projecting the MODWPT coefficients back onto their rescaled filter coefficients via

$$x_t = \sum_{n=0}^{2^j-1} \sum_{l=0}^{L_j-1} \tilde{u}_{j,n,l}\widetilde{w}_{j,n,t+l \ modN}, \quad t = 0, \cdots, N-1, \tag{3.15}$$

where $L_j = (2^j - 1)(L - 1) + 1$ is the length of a level j wavelet filter.

Let $\widetilde{\mathbf{d}}_{j,n} = (\tilde{d}_{j,n,0}, \tilde{d}_{j,n,1,\cdots,\tilde{d}_{j,n,N-1}})$ be the MODWPT detail associated with the frequency interval $\lambda_{j,n}$, then the $t-$th element of $\widetilde{\mathbf{d}}_{j,n}$ is given by

$$\widetilde{d}_{j,n,t} = \sum_{l=0}^{L_j-1} \widetilde{u}_{j,n,l}\widetilde{w}_{j,n,t+l \ mod \ N}, \ t = 0, \cdots, N-1,$$

and an additive decomposition in Equation (3.15) may be rewritten as

$$\mathbf{x} = \sum_{(j,n)\in\mathcal{B}} \widetilde{\mathbf{d}}_{j,n}$$

for any orthonormal basis $\mathcal{B}$. These details are associated with zero-phase filters, just like the MODWT, and therefore line up perfectly with the features in the original time series $\mathbf{x}$ at the same time.

Chapter 4

Estimation Methods for Stationary Long Memory Processes: A Review

Long memory models have been used by several authors to model data with persistent autocorrelations. The earlier model was introduced by Mandelbrot and Van Ness (1968) and Mandelbrot (1971) to formalize Hurst's empirical findings using cumulative river flow data. Let $Y_t = \int_{\infty}^{t} (t-s)^{H-\frac{1}{2}} dB(s)$, where $B(s)$ is Brownian motion and $H \in (0,1)$. Then $X_t = Y_t - Y_{t-1}$ is a simple fractional Gaussian noise, which is designed to account for the long term behavior of time series. The second long memory model was proposed by Granger and Joyeux (1980) and Hosking (1981) – the fractional integrated autoregressive moving average (ARFIMA) models with an infinite peak in the spectrum at $f = 0$. Geweke and Porter-Hudak (1983) proved that the definitions of fractional Gaussian noise and integrated (or fractionally differenced) series are equivalent.

A simple generalization of the ARFIMA model is the Gegenbauer process (Gray et al., 1989) or a seasonal persistent process (Anděl, 1986). More generally, Woodward et al. (1998) proposed the k-factor Gegenbauer and k-factor Gegenbauer ARMA process, which can model long term periodic behavior associated with several peaks in frequencies in $[0, \frac{1}{2}]$.

In order to allow for different persistence parameters across different frequencies, we consider the general class of fractionally integrated zero mean processes $(X_t)_t$, referred to as Seasonal and/or Cyclical Long Memory (SCLM henceforth) processes, defined by the following equation:

$$F(B)X_t = (I-B)^{d_0} \prod_{i=1}^{k-1} (I - 2\nu_i B + B^2)^{d_i} (I+B)^{d_k} X_t = \varepsilon_t, \tag{4.1}$$

where B is the backshift operator. For $i = 1, \cdots, k-1$, $\nu_i = cos\lambda_i$, λ_i being any frequency between 0 and π. For $i = 0, 1, \cdots k$ and d_i is such that: $|d_i| < 1/2$, implying thus that the spectral density is unbounded at λ_i. Moreover, $(\varepsilon_t)_t$ is an innovation process to be specified. This model has been first discussed by Robinson (1994) in order to test

whether the data stems from a stationary or a non-stationary process, under uncorrelated and weakly correlated innovations $(\varepsilon_t)_t$.

Without loss of generality, we assume that $(X_t)_t$ described by Equation $(4.1)$is a zero mean process and for the moment we assume that $(X_t)_t$ is a stationary process. The process nests all the specific long memory processes introduced from the nineties in the literature to take both the seasonal and/or cyclical behaviors and the long memory components of the data into account. From model (4.1), we can derive in a stationary setting a lot of models as follows whose interest in macroeconomics is recognized.

- If $d_i = 0$ $(i = 1, \cdots, k)$, we get the FI(d) (Fractionally Integrated) process if $(\varepsilon_t)_t$ is a white noise:
$$(I - B)^d X_t = \varepsilon_t, \tag{4.2}$$
proposed by Granger and Joyeux (1980) and Hosking (1981). If we assume that $(\varepsilon_t)_t$ follows a GARCH noise, we get the FIGARCH model (fractionally integrated and GARCH), see Baillie, Bollerslev and Mikkelsen (1996) or reference. This class of models permits to take into account the existence of an infinite cycle, as well as the spectral density's typical shape of macroeconomics data, namely an explosion for the very low frequencies.

- In order to model a fixed seasonal periodicity s, supposed to be even, we generally use the following representation:
$$(I - B^s)^d X_t = \varepsilon_t. \tag{4.3}$$
For instance, if $s = 4$, the expression (4.3) becomes:
$$(I - B^4)^d X_t = (I - B)^d (I + B^2)^d (I + B)^d X_t = \varepsilon_t. \tag{4.4}$$
This filter was introduced by Porter-Hudak (1990), called the rigid filter and can be considered as the particular case of Hassler's flexible filter (1994): $(I - B)^{d_1}(I + B^2)^{d_2}(I + B)^{d_3} X_t = \varepsilon_t$. It is motivated by factorizing $I - B^4$ according to its unit roots, allowing to model stationary fractional seasonalities. The spectral density of the process (4.3) is unbounded for the three seasonal frequencies $\{0, \pi/2, \pi\}$. This representation is useful for quarterly data sets (with $s = 4$) and if we deal with monthly data we consider the same model using $s = 12$.

- It may happen that we observe an explosion at the zero frequency as well as at any frequency between $]0, \pi]$. This means that an infinite cycle is mixed with another seasonality. In that case, we use the following model for $(X_t)_t$:
$$(I - B)^{d_1}(I - B^s)^{d_2} X_t = \varepsilon_t. \tag{4.5}$$
This model was introduced by Porter-Hudak (1990). The parameter d_1 corresponds to the persistence associated to the infinite cycle and d_2 is the persistence associated to the fixed seasonality.

- In the presence of explosions at k frequencies in the spectral density, we use a model characterizing these k persistent periodicities. It is the k-factor Gegenbauer process given by:

$$\prod_{i=1}^{k}(I - 2\nu_i B + B^2)^{d_i} X_t = \varepsilon_t, \tag{4.6}$$

  with, for $i = 1, \ldots, k$, $\nu_i = cos(\lambda_i)$, the frequencies $\lambda_i = \cos^{-1}(\nu_i)$ being the Gegenbauer frequencies or the G- frequencies. When $(\varepsilon_t)_t$ is a white noise, this model was introduced by Giraitis and Leipus (1995) and Woodward, Cheng and Gray (1998). When $(\varepsilon_t)_t$ follows a GARCH process, $(X_t)_t$ is called a GIGARCH process, introduced by Guégan (2000, 2003).

- Inside the spectral density, we can observe k explosions as well as an explosion at the zero frequency, then the previous model becomes:

$$(I - B)^d \prod_{i=1}^{k}(I - 2\nu_i B + B^2)^{d_i} X_t = \varepsilon_t. \tag{4.7}$$

There are a number of estimators of a long memory process' long memory parameter when the parameter is assumed to hold constant over the entire data set. We make a review in the following.

4.1 ARFIMA Processes

A model that has long range dependence and is frequently used in modeling long memory is the fractionally integrated autoregressive moving average (ARFIMA) model. The ARFIMA model succinctly captures the slowly decaying autocovariance function of a long memory process with fractional difference or long memory parameter $d = H - \frac{1}{2}$, where H is the Hurst parameter. By letting the difference parameter take non-integer values, the ARFIMA model permits to model complex long-run behavior in a more parsimonious manner.

There are several estimation methods for ARFIMA models, including parametric, semiparametric (in frequency domain and in time domain), and wavelet methods. Nielsen and Frederiksen (2005) studied and carried out the finite sample comparison of different estimation methods. They found that among the parametric method, the frequency domain maximum likelihood procedure is superior with respect to both bias and RMSE. And the bias of parametric time domain procedures is alleviated when larger sample sizes are considered. Among the semiparametric (frequency domain and wavelet) methods, the bias-reduced log-periodogram regression (Andrews and Guggenberger, 2003) and local polynomial Whittle estimators (Andrews and Sun, 2004) outperform the correctly specified time domain parametric domain methods. Furthermore, when the methods are very heavily biased owing to contamination from short-run dynamics, these estimators show a

much lower bias at the expense of an increase in the RMSE. Finally, if there is not sufficient trimming of scales, the wavelet-based method are heavily biased when short-run dynamics is introduced.

We first briefly describe the ARFIMA model and provide an introduction to the estimation methods. We do not present all the mathematical assumptions underlying each estimation procedure but rather describe the methods and applicability in general and also briefly discuss and compare the asymptotic distributions of the various estimators.

Granger and Joyeux (1980) and Hosking (1981) introduced an $ARFIMA(p, d, q)$ process if its d-th difference is a stationary and invertible $ARMA(p, q)$ process. Here d may be any real number such that $-\frac{1}{2} < d < \frac{1}{2}$ (to ensure the stationarity and invertibility). More precisely, X_t is an $ARFIMA(p, d, q)$ if

$$\Phi(B)(I - B)^d(X_t - \nu) = \Theta(B)\varepsilon_t, \tag{4.8}$$

where $\Phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$ and $\Theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$ are polynomials of order p and q, respectively, with roots outside the unit circle, ε_t is Gaussian white noise with the variance σ^2.

If $d > -\frac{1}{2}$, the process is invertible and possesses a linear (Wold) representation, and if $d < \frac{1}{2}$, it is covariance stationary. If $d = 0$, the spectral density is bounded at the origin and the process has only weak dependence (short memory). Furthermore, the parameter d determines the (long) memory of the process. If $d > 0$, the process is said to have long memory, since the autocorrelations die out at a hyperbolic rate (and indeed are no longer absolutely summable), in contrast to the much faster exponential rate in the weak dependence case, whereas if $d < 0$, the process is said to be antipersistent (Mandelbrot, 1982) and has mostly negative autocorrelations. The case $0 \leq d \leq \frac{1}{2}$ has been proved particularly relevant for many applications in finance and economics.

The autocorrelation function of the process (4.8) satisfies

$$\rho_k \sim c_\rho k^{2d-1}, \;\; 0 < c_\rho < \infty \;\; as \; k \to \infty, \tag{4.9}$$

which decays at a hyperbolic rate, see Granger and Joyeux (1980) and Hosking (1981). Equivalently, the behavior of the autocorrelations at large lags can be stated in the frequency domain at small frequencies.

Thus, defining the spectral density function of X_t, $f_X(\lambda)$, as

$$\gamma_k = \int_{-\pi}^{\pi} f_X(\lambda) e^{i\lambda_k} d\lambda, \tag{4.10}$$

where γ_k is the k-th autocovariance of X_t, it can be shown that the spectral density of the $ARFIMA(p,d,q)$ process (4.8) is given by

$$f_X(\lambda) = \frac{\sigma^2}{2\pi}|1-e^{i\lambda}|^{-2d}\frac{|\Theta(e^{i\lambda})|^2}{|\Phi(e^{i\lambda})|^2} = \frac{\sigma^2}{2\pi}(2\sin\lambda/2)^{-2d}\frac{|\Theta(e^{i\lambda})|^2}{|\Phi(e^{i\lambda})|^2}. \tag{4.11}$$

Thus the approximation (4.9) can be restated in the frequency domain as follows (see Granger and Joyeux (1980), Hosking (1981), or Beran (1994))

$$f_X(\lambda) \sim g|\lambda|^{-2d}, \;\; 0 < g < \infty, \;\; as\ \lambda \to 0. \tag{4.12}$$

4.1.1 Parametric Estimators

Four different parametric maximum likelihood estimators (MLE) are described in the following: the exact time domain MLE, the modified profile likelihood estimator, the conditional time domain MLE, and the frequency domain MLE. The time domain estimators are based on the likelihood function of the ARFIMA(p,d,q) model with or without conditioning on initial observations, and the frequency domain estimator is based on Whittle's approximation to the likelihood function in the frequency domain.

Maximum Likelihood in the Time Domain

As for the estimation of the factional differencing parameter in ARFIMA model, in time domain, the often used maximum likelihood method are the exact Gaussian maximum likelihood estimator (Sowell, 1992b), the modified likelihood estimator (Cox and Reid, 1987; An and Bloomfield, 1993), the conditional maximum likelihood estimator or conditional sum of squares estimator (Chung et Baillie, 1993; Beran, 1995; Tanaka, 1999; Nielsen, 2004), etc.

The exact Gaussian maximum likelihood objective function for the model (4.8) is (when $-\frac{1}{2} < d < \frac{1}{2}$):

$$L_E(d,\phi,\theta,\sigma^2,\mu) = -\frac{T}{2}\log|\Omega| - \frac{1}{2}(Y-\mu l)'\Omega^{-1}(Y-\mu l), \tag{4.13}$$

where $l = (1,\cdots,1)'$, $X = (X_1,\cdots,X_T)'$, ϕ and θ are the parameters of $\Phi(B)$ and $\Theta(B)$, μ is the mean of X, and Ω is the variance matrix of X, which is a complicated function of d and the remaining parameters of the model. Sowell (1992a) derived an efficient procedure for solving this function in terms of hypergeometric functions, but an important limitation is that the roots of the autoregressive polynomial cannot be multiple.

Gathering the parameters in the vector $\gamma = (d,\phi,\theta,\sigma^2,\mu)'$, the exact maximum likelihood (EML) estimator is obtained by maximizing the likelihood function (4.13) with respect to γ. Sowell (1992a) showed that the EML estimator of d is $\sqrt{T}$-consistent and asymptotically normal, i.e.,

$$\sqrt{T}(\hat{d}_{EML} - d) \to^d N(0,(\pi^2/6 - C)^{-1}), \tag{4.14}$$

where $C = 0$ when $p = q = 0$ and $C > 0$ otherwise. The variance of the EML estimator may be derived as the $(1, 1)$ element of the inverse of the matrix

$$\frac{1}{4\pi}\int_0^{2\pi}\frac{\partial \log f_X(\lambda)}{\partial \gamma}\frac{\partial \log f_X(\lambda)}{\partial \lambda}d\lambda.$$

Although the time and frequency domain maximum likelihood estimators are asymptotically equivalent, their finite sample properties differ. Sowell (1992b) showed that the time domain estimator has better finite sample properties than the frequency domain estimator when the mean of the process is known. However, Cheung and Diebold (1994) showed that the finite sample efficiency of the discrete Whittle frequency domain MLE relative to time domain EML rised dramatically when the mean is unknown and has to be estimated.

The modified profile likelihood (MPL) estimator is based on a correction of the parameters of interest (here d, Φ, Θ) for the second-order effects due to nuisance parameters (here σ^2, μ). Thus the idea is to reduce the bias by applying a transformation that makes (d, Φ, Θ) orthogonal to (σ^2, μ), see Cox and Reid (1987) and An and Bloomfield (1993). The modified profile log-likelihood function is given as (without constants)

$$L_M(d, \phi, \theta; \hat{\mu}) = -(\frac{1}{2}-\frac{1}{T})\log|R| - \frac{1}{2}\log(l'R^{-1}l) - (\frac{T-3}{2})\log[T^{-1}(X-\hat{\mu}l)'R^{-1}(X-\hat{\mu}l)], \tag{4.15}$$

where $R = \Omega/\sigma^2$ and $\hat{\mu} = (l'R^{-1}l)^{-1}l'R^{-1}X$. The asymptotic distribution of the MPL estimator is unchanged compared to the EML estimator on which it is based, and hence it also satisfies the asymptotic distribution (4.14).

Imposing the initialization $X_t = 0 (t \leq 0)$, the model (4.8) is valid for any value of d and is a type II fractional process in the terminology of Marinucci and Robinson (1999). The objective function considered by Chung and Baillie (1993), Beran (1995), Tanaka (1999), and Nielsen (2004) is

$$L_c(d, \phi, \theta, \mu) = -\frac{T}{2}\log[\sum_{t=1}^{T}(\frac{\Phi(B)}{\Theta(B)}(I-B)^d(X_t-\mu))^2], \tag{4.16}$$

and we call the estimator that maximizes Equation (4.16) the conditional maximum likelihood (CML) estimator. Maximizing L_C is equivalent to minimizing the usual (conditional) sum of squares, and hence this estimator is also referred to as the CSS estimator by some authors, e.g., Chung and Baillie (1993) and Beran (1995). The CML estimator has the same asymptotic distribution (4.14) as the EML estimator for any value of d and is computationally much less demanding.

The parametric estimators are asymptotically efficient in the classical sense when the model is Gaussian and correctly specified.

Maximum Likelihood in the Frequency Domain

An alternative approximate MLE of the $ARFIMA(p, d, q)$ model follows the idea of Whittle (1951), who noted that for stationary models the covariance matrix Ω can be

diagonalized by transforming the model into the frequency domain. Fox and Taqqu (1986) showed that (when $d \in (-\frac{1}{2}, \frac{1}{2})$) the log likelihood can be approximated by

$$L_F(d, \phi, \theta, \sigma^2) = -\sum_{j=1}^{[T/2]} [\log f_X(\lambda_j) + \frac{I(\lambda_j)}{f_X(\lambda_j)}], \tag{4.17}$$

where $\lambda_j = 2\pi j/T$ are the Fourier frequencies, $I(\lambda) = \frac{1}{2\pi T}|\sum_{t=1}^{T} X_t e^{it\lambda}|^2$ is the periodogram of X_t, $f_X(\lambda)$ is the spectral density of X_t given in Equation (4.11).

The approximate frequency domain maximum likelihood (FML) estimator is defined as the maximizer of Equation (4.17) and was proposed by Fox and Taqqu (1986), who also proposed a continuously integrated version of (4.17). Dahlhaus (1989) also assumed Gaussianity and considered the exact likelihood function in the frequency domain. Note that the FML estimator is invariant to the presence of a nonzero mean, i.e., $\mu \neq 0$, since $j = 0$ (the zero frequency) is left out of the summation in (4.17). The FML estimator has the same asymptotic normality, i.e., $\sqrt{T}$-consistency and asymptotic normality, and when the process is Gaussian, asymptotic efficiency. Finally, Giraitis and Surgailis (1990) relaxed the Gaussianity assumption and analyze the Whittle estimate for linear processes, showing that it is $\sqrt{T}$-consistent and asymptotically normal but no longer efficient, while Hosoya (1997) extended the previous analysis to multivariate framework.

4.1.2 Semiparametric Estimators

After the seminal papers of Granger and Joyeux (1980) and Hosking (1981), fractional integration processes (FI(d)) have attracted the attention of many statisticians and econometricians. These long-range dependence processes give more flexibility to empirical research than the classical FI(0) and FI(1) classes of processes. For $0 < d < 1/2$, they are stationary with hyperbolic decay of the autocorrelation function and they exhibit long memory or long-range dependence. To estimate d, we usually use the semiparametric methods developed by Geweke and Porter-Hudak (1983) (henceforth referred to GPH). Agiakloglou et al. (1993) showed that this estimator has a large bias. Reisen (1994), Robinson (1994, 1995a,b) and Lobato and Robinson (1996) give some estimators with a small bias.

The semiparametric frequency domain estimators are based on the approximation (4.12) to the spectral density. Two classes of semiparametric estimators have become very popular in empirical work, the log-periodogram regression method suggested by Geweke and Porter-Hudak (1983) and the local Whittle approach suggested by Künsch (1987). Some earlier work on the (adjusted) rescaled range, or "R/S" statistic, by Hurst (1951) and Mandelbrot and Wallis (1969), or its modified version allows for weak dependence by Lo (1991).

The semiparametric estimators enjoy robustness to short-run dynamics, since they use

only information from the periodogram ordinates in the vicinity of the origin. Indeed, the short-run dynamics in the model, i.e., the autoregressive and moving average polynomials $\Phi(\cdot)$ and $\Theta(\cdot)$ in ARFIMA model (4.8), does not even have to be specified. The drawback is that only $\sqrt{m}$-consistency is achieved, where $m = m(T)$ is a user-chosen bandwidth parameter, in comparison to $\sqrt{T}$-consistency (and efficiency) in the parametric case. Thus the semiparametric approach is much less efficient than the parametric one, sine it requires at least $\frac{m}{T} \to 0$.

Log-periodogram Regression

Probably the most commonly applied semiparametric estimator is the log-periodogram regression (LPR) estimator introduced by Geweke and Porter-Hudak (1983) and analyzed in detail by Robinson (1995b). Taking logs in Equation (4.12) and inserting sample quantities, we get the approximate regression relationship

$$\log(I(\lambda_j)) = constant - 2d\log(\lambda_j) + error. \tag{4.18}$$

The LPR estimator is defined as the OLS estimator in the regression (4.18) using $j = 1, \cdots, m$, where $m = m(T)$ is a bandwidth number which tends to infinity as $T \to \infty$ but at a slower rate than T. Note that the estimator is invariant to a non-zero mean, since $j = 0$ is left out of the regression.

Under suitable regularity conditions, including X_t being Gaussian (later relaxed by Velasco (2000)) and a restriction on the bandwidth, Robinson (1995b) derived the asymptotically normal limit distribution for the LPR estimator when d is in the stationary and invertible range $(-\frac{1}{2}, \frac{1}{2})$. The proof by Robinson (1995b) also employed trimming of the very low frequencies as suggested by Künsch (1986), but following recent research, e.g., Hurvich et al. (1998) and the original suggestion of Geweke ad Porter-Hudak (1983), the trimming is not necessary and has been largely ignored in empirical work. Kim and Phillips (1999) and Velasco (1999b) demonstrated that the range of consistency is $d \in (-\frac{1}{2}, 1]$ and the range of asymptotic normality is $d \in (-\frac{1}{2}, \frac{3}{4})$. The limiting distribution of the LPR estimator for $d \in (-\frac{1}{2}, \frac{1}{2})$ is given by Robinson (1995a) as

$$\sqrt{m}(\widehat{d}_R - d) \to^d N(0, \frac{\pi^2}{24}). \tag{4.19}$$

There exist also some other estimators based on LPR estimator, the bias reduced log-periodogram regression (BRLPR) estimator by Agiakloglou et al. (1993) and Andrews and Gugggenberger (2003) and the pooled log-periodogram regression (PLPR) estimator by Shimotsu and Phillips (2002); see Nielsen and Frederisksen (2005) for more details.

Local Whittle Approach

The other class of semiparametric frequency domain estimators we consider follows the local Whittle approach suggestd by Künsch (1987). The local Whittle (LW) estimator was

analyzed by Robinson (1995a) (who called it a Gaussian semiparametric estimator) and is attractive because of its likelihood interpretation, nice asymptotic properties and very mild assumptions. The LW estimator is defined as the maximizer of the (local Whittle likelihood) function

$$\mathcal{Q}(g,d) = -\frac{1}{m}\sum_{j=1}^{m}[\log(g\lambda_j^{-2d}) + \frac{I(\lambda_j)}{g\lambda_j^{-2d}}]. \tag{4.20}$$

One drawback compared to log-periodogram estimation is that numerical optimization is needed. However, the assumptions underlying this estimator are weaker than those of the LPR estimator, and Robinson (1995a) showed that when $d \in (-\frac{1}{2}, \frac{1}{2})$,

$$\sqrt{m}(\widehat{d}_{LW} - d) \rightarrow^d N(0, \frac{1}{4}). \tag{4.21}$$

Thus the asymptotic distribution is extremely simple, facilitating easy asymptotic inference, and in particular the estimator is more efficient than the LPR estimator. The ranges of consistency and asymptotic normality for the LW estimator have been shown by Velasco (1999a) and by Phillips and Shimotsu (2004) to be the same as those of the LPR estimator.

An exact local Whittle (ELW) estimator has been proposed by Shimotsu and Phillips (2005) that avoids some of the approximations in the deviation of the LW estimator and is valid for any value of d. The ELW estimator replaces the objective function (4.20) by the function

$$\mathcal{Q}_E(g,d) = -\frac{1}{m}\sum_{j=1}^{m}[\log(g\lambda_j^{-2d}) + \frac{I_{\Delta^d X}(\lambda_j)}{g}], \tag{4.22}$$

where $I_{\Delta^d X}(\lambda) = \frac{1}{2\pi T}|\sum_{t=1}^{T}(\Delta^d X_t)e^{it\lambda}|^2$ is the periodogram of $\Delta^d X_t$. The ELW estimator satisfies (4.21) for any value of d and is thus not confined to any particular range of d values, but it is confined to zero-mean processes.

In addition, Shimotsu (2002) proposed the feasible ELW (FELW) estimator. Andrews and Sun (2004) proposed a generalization of the local Whittle estimator–the local polynomial Whittle (LPW) estimator of d for $d \in (-\frac{1}{2}, \frac{1}{2})$.

For both the log-periodogram regression method and the local Whittle approach, the choice of bandwidth parameter m is very important. Results on optimal (mean squared error minimizing) choice of bandwidth for the log-periodogram regression have been derived by Hurvich et al. (1998), and results for the local Whittle approach have been derived by Henry and Robinson (1996). In both cases, the optimal bandwidth is found to be a multiple of $T^{0.8}$, where the multiplicative constant depends on the smoothness of the spectral density near the origin, i.e., on the short-run dynamics of the process.

4.1.3 Wavelet Estimators

In the latter part of the twenties century, several disciplines came to the realization of the naturally occurring phenomena (river flow, atmospheric patterns, telecommunications, astronomy, financial markets, etc.) which exhibit correlations that do not decay at a sufficiently fast rate. This means that observations separated by long periods of time still exhibit significant correlation. These time series are known as long memory or long range dependent processes and require different approaches to modeling than short memory time series. At first, Fourier-based methods dominated the literature in terms of identifying and fitting models to long memory processes. Both least squares and maximum likelihood procedures have been established for estimating the model parameters in the case of long range dependence. By generalizing the concept of a long memory process to a long memory process with seasonality, one may apply these estimation procedures to a much broader class of time series models.

Wavelets have shown great promise in handling long memory and a combination of short-memory and long-memory processes. By performing a wavelet decomposition to an observed time series, one makes an implicit assumption about the underlying nature of the process. Specifically, there is a hierarchical structure present in the data so that information in the time series is evolving at different time horizons (scales) and different magnitudes. When decomposed appropriately, one may more easily view the individual factors that make up the (potentially) complicated process. More precisely, the correlation matrix of the decomposition is essentially block diagonal. Large values on the diagonal represent the information at a particular level of the decomposition and near zero values on the off-diagonal denote little interaction between elements. This concept was discussed by Brock (2000).

By design, the wavelet's strength rests in its ability to localize simultaneously processes in time and scale. At high scales, the wavelet has a small centralized time support enabling it to focus on short-lived time phenomena like a singularity point. At low scales, the wavelet has a large time support allowing it to identify long periodic behavior. By moving from low to high scales, the wavelet zooms in on the behavior of a process at a particular point in time, identifying singularities, jumps, and cusps. Alternatively, the wavelet can zoom out to reveal the long, smooth features of a series. In practice, the Haar and Daubechies (1988) wavelets are most commonly applied in the literature.

In the study of the estimation of long memory parameters, there exists mainly two methods through wavelet techniques, the ordinary least squares (OLS) method and the maximum likelihood method. These methods are applicable due to the ability of the DWT to decorrelate long memory processes.

For the fractional integrated ARMA (ARFIMA) model introduced by Granger and Joyeux (1980) and Hosking (1981), an explosion can be found at the zero frequency in its spectrum. Using the expression of the spectral density of the ARFIMA process, Geweke and Porter-Hudak (1983) proposed a semiparametric method using the log regression on the

periodogram in order to estimate the long memory parameter d. We refer to this estimator the GPH estimator. Thereafter, several improvements have been suggested such as using an alternative spectral estimator. In the wavelet framework, Jensen (1999) introduced a wavelet-based estimator of d using the OLS regression. The main technique used here is that the wavelet coefficients' variance is an approximate estimation of the spectrum. Thus we can obtain a log linear regression between the wavelet coefficients' variance and the scale. This approach is known for its simplicity in operation and application, but the large variance of the estimator posed a problem. Thus we need to pay attention for the adoption of this method.

The other method often used by the practitioners is the wavelet-based approximate maximum likelihood estimation, which has been investigated by McCoy and Walden (1996) and Jensen (1999a, 2000). This method produced much less mean squared errors when compared to the wavelet-based OLS method. To avoid the complexity in computing the exact likelihood, a wavelet-based approximate maximum likelihood estimator is proposed by replacing the covariance matrix of the process with an approximation using the DWT.

Let X_t be a stationary and invertible FI(d) ($-\frac{1}{2} < d < \frac{1}{2}$) process with zero mean whose d-th order backward difference is as follows:

$$(I-B)^d X_t = \sum_{k=0}^{\infty} \begin{pmatrix} d \\ k \end{pmatrix} (-1)^k X_{t-k} = \varepsilon_t, \tag{4.23}$$

where $\begin{pmatrix} d \\ k \end{pmatrix} = \frac{d!}{k!(d-k)!}$ and ε_t is a white noise process with variance σ_ε^2.

The autocovariance sequence (ACVS) of X_t is defined to be

$$\gamma_\tau = E(X_t, X_{t-\tau}) = \frac{\sigma_\varepsilon^2 (-1)^\tau \Gamma(1-2d)}{\Gamma(1+\tau-d)\Gamma(1-\tau-d)},$$

which means the variance is given by

$$Var(X_t) = \gamma_0 = \frac{\sigma_\varepsilon^2 \Gamma(1-2d)}{[\Gamma(1-d)^2]}.$$

The spectral density function (SDF) of X_t is

$$S_X(f) = \frac{\sigma_\varepsilon^2}{|2\sin(\pi f)|^{2d}}, \quad for \quad -\frac{1}{2} < f < \frac{1}{2}. \tag{4.24}$$

Consequently, $S_X(f) \propto f^{-2d}$ approximately as $f \to 0$ and, thus the SDF is approximately linear on the log scale. When $0 < d < \frac{1}{2}$, the SDF has an asymptote at frequency zero, in which case the process exhibits slowly decaying autocovariances and provides a simple example of a long memory process.

We explore the output of the DWT when applied to an FI(d) process. Emphasis will be on the spectral properties of the DWT coefficients. The ability of the DWT to decorrelate time series, such as the FI(d) process, producing wavelet coefficients for a given scale, which are approximately uncorrelated, is well-known; see, for example, Tewfik and Kim (1992), McCoy and Walden (1996) and Wornell (1996). The DWT produces wavelet coefficients with flat spectra and therefore exhibit little correlation within scales. And there is very little cross-correlation between scales too. This property is used to simulate FI(d) process and also to estimate the long memory parameter d in practice. Ramsey (1998) showed this fact empirically to validate previous results with respect to performing regressions at different time scales.

There is an important property worthy of being mentioned. The band-pass variance B_j for an FI(d) process, with spectrum given in Equation (4.24), in the frequency interval $[-2^{-j}, -2^{-j-1}) \cup (2^{-j-1}, 2^{-j}]$ is

$$B_j = 2 \cdot 4^{-d} \int_{1/2^{j+1}}^{1/2^j} \frac{\sigma_\varepsilon^2}{|\sin(\pi f)|^{2d}} df. \tag{4.25}$$

Replace the true SDF at each frequency interval with a constant $S_j = S_j(f)$, for all f, such that the band-pass variances are equal. This step assumes the SDF is slowly varying across the frequency interval. Integrating the constant spectrum over $[-2^{-j}, -2^{-j-1}) \cup (2^{-j-1}, 2^{-j}]$ gives

$$\int_{1/2^{j+1}}^{1/2^j} S_{j,n} df = S_j 2^{-j-1}.$$

Equating this to the band-pass variance gives

$$2S_j 2^{-j-1} = B_j \implies S_j = 2^j B_j. \tag{4.26}$$

The variance of wavelet coefficients $w_{j,t}$ is therefore given by S_j through the band-pass nature of the DWT.

A popular time-domain technique for simulating time series is based on the Levinson-Durbin recursions (Hosking, 1984). This time domain method is exact, but the Levinson-Durbin recursions require $O(N^2)$ operations and become unwieldy for larger sample sizes. Davies and Harte (1987) described a method for simulating certain stationary Gaussian time series of length N with known autocovariances $\gamma_0, \gamma_1, \cdots, \gamma_{N-1}$. The method is based on the Fourier transform and is provided in Chan and Wood (1994) and Beran (1994). This method is exact for short memory processes, like the typical ARMA time series models, but is approximate when used to generate FI(d) process (Percival, 1993). Even so, the Fourier-based method is very efficient and produces realizations of FI(d) process with good statistical properties. Recently, Parke (1999) proposed a simulation procedure by representing the FI(d) process via an error duration model. A DWT-based method for generating realizations of FI(d) process was proposed by McCoy and Walden (1996).

Wavelet Ordinary Least Squares (OLS) Estimator

Using the logarithmic decay of the autocovariance function of a long memory process, Jensen (1999) showed that a log-linear relationship (suggested by McCoy and Walden, 1996; Johnstone and Silverman, 1997) exists between the variance of the wavelet coefficient from the long memory process and its scale, which can be used to estimate d by least squares regression. Leaving out high level wavelet coefficients results in robustness to the short-run dynamics, which is similar to the LPR estimator (see McCoy and Walden, 1996; Tse et al., 2002).

The approximate linear relationship between the periodogram and Fourier frequencies (on a log-log scale) has been known for a long time. Geweke and Porter-Hudak (1983) first proposed the regressing the log values on the periodogram on the log SDF to estimate the fractional differencing parameter d (we refer to this as the GPH estimator). Although very popular, the GPH estimator suffers from the poor asymptotic properties of the periodogram. Improvements to the GPH estimator have been suggested, such as restricting the number of frequencies used in the regression or using an alternative spectral estimator (smoothed or multitaper). A wavelet-based estimator of d was introduced by Jensen (1999b) using ordinary least square (OLS) regression. Tkacz (2000) has applied this technique to nominal interest rates in the United States and Canada.

For a vector of observations $\mathbf{y}$, the OLS model is formulated via $\mathbf{y} = \mathbf{X}\beta + \mathbf{e}$, where

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_J \end{pmatrix} \quad \mathbf{X} = \begin{pmatrix} 1 & x_{1,1} \\ 1 & x_{2,1} \\ \vdots & \vdots \\ 1 & x_{J,1} \end{pmatrix} \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$$

are the length J vector of dependence observations, the $J \times 2$ dimensional model matrix and the length 2 parameter vector, respectively. The final vector

$$\mathbf{e} = [e_1, e_2, \cdots, e_J]^T$$

is the column of model errors with $E(\mathbf{e}) = 0$ and $Var(\mathbf{e}) = \sigma_e^2 \mathbf{I}_J$. The OLS estimator $\widehat{\beta}$ of β is

$$\widehat{\beta} = (\mathbf{X^T X})^{-1} \mathbf{X^T y}, \tag{4.27}$$

and the covariance matrix of $\widehat{\beta}$ is given by

$$\Sigma_{\widehat{\beta}} = \sigma_e^2 (\mathbf{X^T X})^{-1}. \tag{4.28}$$

We are interested in estimating the slope parameter β_1 when the dependent observations are $y_j = \log(\widehat{\sigma}_x^2(\lambda_j))$ and the independent observations are $x_{j,1} = \log(\lambda_j)$, for $j = 1, \cdots, J$. There is a relationship between the wavelet variance $\sigma_x^2(\lambda_j)$ and scale λ_j (on a log-log scale) for FI(d) process. Jensen (1999b) proved that $\sigma_x^2(\lambda_j) \to C\lambda_j^{2d-1}$ as $j \to \infty$ for FI(d) process with $-\frac{1}{2} \leq d \leq \frac{1}{2}$. A reasonable regression model is therefore

$$\log(\hat{\sigma}_x^2(\lambda_j)) = \beta_0 + \beta_1 \log(\lambda_j) + e_j, \tag{4.29}$$

where $\beta_1 = 2d - 1$. Defining $y_j = \log(\hat{\sigma}_x^2(\lambda_j))$ and $x_{j,1} = \log(\lambda_j)$ for $j = 1, \cdots, J$, the OLS estimator of β_1 is given by the second element of $\widehat{\beta}$ in Equation (4.27), or explicitly,

$$\widehat{\beta}_1 = \frac{\sum_{j=1}^{J}[\log(\lambda_j) - \overline{\log(\lambda_j)}]\log(\hat{\sigma}_x^2(\lambda_j))}{\sum_{j=1}^{J}[\log(\lambda_j) - \overline{\log(\lambda_j)}]^2},$$

where $\overline{\log(\lambda_j)}$ is the sample mean of $\log(\lambda_j)$. Hence the OLS estimator of the fractional difference parameter is $\hat{d} = (\hat{\beta}_1 + 1)/2$. The variance of the OLS estimator is

$$Var(\hat{\beta}_1) = \frac{\hat{\sigma}_e^2}{\sum_{j=1}^{J}[\log(\lambda_j) - \overline{\log(\lambda_j)}]^2},$$

where the estimated variance of the model errors is given by (most easily expressed in matrix notation)

$$\widehat{\sigma}_e^2 = \frac{(y - X\widehat{\beta})^T(y - X\hat{\beta})}{J - 2}.$$

Basic properties of the variance tell us that $Var(\hat{d}) = \frac{1}{4}Var(\hat{\beta}_1)$. Jensen (2000) used the wavelet transform to decompose the variance of a long memory process to develop an alternative to the frequency domain estimators of the long memory parameter, but only for globally stationary long memory processes.

In particular, Jensen (1999) showed that for the fractionally integrated noise process X_t satisfying $(I - B)^d X_t = \varepsilon_t$, when $d \in (-\frac{1}{2}, \frac{1}{2})$, the wavelet coefficient $\omega_{j,k}$ has the following asymptotic behavior,

$$\omega_{j,k} \rightarrow^d N(0, \sigma^2 2^{-2jd}) \quad as\ j \rightarrow 0. \tag{4.30}$$

If we define the variance of $\omega_{j,k}$ as $R(j)$, the intuitive log-linear relationship

$$\log R(j) = \log \sigma^2 - d \log 2^{2j} \tag{4.31}$$

arises. To estimate d through Equation (4.31), an estimate of the variance is required. Jensen (1999) proposed

$$\widehat{R}(j) = 2^{-j} \sum_{k=0}^{2^j - 1} \omega_{j,k}^2, \quad j = 0, \cdots, p - 1, \tag{4.32}$$

and the relationship (4.31) thus gives rise to the regression

$$\log \widehat{R}(j) = constant - d \log 2^{2j} + error, \quad j = J, \cdots, p - 1 - K, \tag{4.33}$$

which can be estimated by ordinary least squares yielding the wavelet OLS (WOLS) estimator. The WOLS estimator is consistent and asymptotically normal when $d \in (-\frac{1}{2}, \frac{1}{2})$, see Jensen (1999). The trimming of the lowest J scales was suggested by Jensen (1999)

to avoid boundary effects, and the trimming of the highest K scales was suggested by McCoy and Walden (1996) and Tse et al. (2002), since (4.30) is valid for small j only.

The wavelet coefficients' variance is a regularization of the spectrum (Percival, 1995, McCoy and Walden, 1996), which is a quite useful property for the estimation theory. The wavelet coefficients' variance decomposes the variance of the series across different scales. Those scales which contribute the most to the series' variance are associated with those wavelet coefficients with largest variance. Hence, the wavelet coefficients' sample variance provided a more intuitive parametric estimate of its population variance than the non-parametric periodogram does of the power spectrum. More importantly, whereas the periodogram is an inconsistent estimator of the spectrum, the wavelet coefficients' sample variance is a consistent estimator of the population variance that enables the wavelet OLS estimator to be a consistent estimator of the fractional differencing parameter.

The use of periodogram regressions appears simple but rather ad hoc. Nevertheless, Yajima (1985) extended these results to show the strong consistency of this least squares estimator over d in the open interval $(-\frac{1}{2}, \frac{1}{2})$ and the asymptotic normality of these least squares estimator for d in $[0, \frac{1}{4})$. In addition, he gave results on the rates of convergence for the estimator when $d \in [\frac{1}{4}, \frac{1}{2})$. The main problem of the least squares technique is its large variance.

Maximum Likelihood Estimation in Scale and Space (Wavelet MLE)

Other researchers (e.g. Hipel and McLeod, 1978; Hosking, 1981) suggested using maximum likelihood (ML) techniques on the basis of asymptotic efficiency. Nevertheless, it is unclear how the long-term and short-term parameters can be distinguished (i.e. identified) in the time domain. Identification can be done very simply in the frequency domain, despite the general equivalence that exists between spectral and time domain frameworks. Hosking (1981) proposed an iterative ML technique that first estimated d using ML techniques and then identified the ARMA structure from a residual series differenced by d. Clearly this estimate of d would be inconsistent.

Wavelet-based maximum likelihood estimation procedure has been investigated by McCoy and Walden (1996) and Jensen (1999a, 2000). Although ordinary least squares estimation is popular because of its simplicity to program and to compute, it produces much larger mean square errors when compared to maximum likelihood method. The wavelet-based maximum likelihood method presented here overcomes the difficulty of computing the exact likelihood by replacing the covariance matrix of the process with an approximation using the DWT. This is possible due to the ability of the DWT to decorrelate long memory processes.

More concretely, let $\mathbf{x}$ be a length $N(= 2^J)$ FI(d) process with mean zero and covari-

ance matrix given by $\Sigma_{\mathbf{x}}$. Then its likelihood can be written as

$$L(d, \sigma_\varepsilon^2|\mathbf{x}) = (2\pi)^{-N/2}|\Sigma_{\mathbf{x}}|^{-1/2}\exp[-\frac{1}{2}\mathbf{x}^T\Sigma_{\mathbf{x}}^{-1}\mathbf{x}], \tag{4.34}$$

see Brockwell and Davis (1991) for reference. The quantity $|\Sigma_{\mathbf{x}}|$ is the determinant of $\Sigma_{\mathbf{x}}$. The maximum likelihood (MLE) estimators of the parameters (d and σ_ε^2) are those quantities that maximize Equation (4.34). To avoid the difficulties in computing the exact MLE, we use the approximate decorrelation of the DWT as applied to FI(d) process; that is ,

$$\Sigma_{\mathbf{x}} \approx \widehat{\Sigma}_{\mathbf{x}} = \mathcal{W}^T\Omega_N\mathcal{W},$$

where $\mathcal{W}$ is the orthonormal matrix defining the DWT and Ω_N is a diagonal matrix containing the variances of DWT coefficients computed from FI(d) process ; that is,

$$\Omega_N = diag(\underbrace{S_1, \cdots, S_1}_{N/2}, \underbrace{S_2, \cdots, S_2}_{N/4}, \cdots, \underbrace{S_j, \cdots, S_j}_{N/2^j}, \cdots, S_J, S_{J+1}).$$

The approximate likelihood function is now

$$\hat{L}(d, \sigma_\varepsilon^2|\mathbf{x}) = (2\pi)^{-N/2}|\widehat{\Sigma}_{\mathbf{x}}|^{-1/2}\exp[-\frac{1}{2}\mathbf{x}^T\widehat{\Sigma}_{\mathbf{x}}^{-1}\mathbf{x}].$$

Hence, we want to find values of d and σ_ε^2 that minimize the log-likelihood function

$$\widehat{\mathcal{L}}(d, \sigma_\varepsilon^2|\mathbf{x}) = -2\log(L(d, \sigma_\varepsilon^2|\mathbf{x})) - N\log(2\pi) = \log(|\widehat{\Sigma}_{\mathbf{x}}|) + \mathbf{x}^T\widehat{\Sigma}_{\mathbf{x}}^{-1}\mathbf{x}.$$

Since the variance for scale λ_j DWT coefficients is given by S_j. We note that S_j depends on two parameters related to the FI(d) process, the fractional differencing parameter d and variance σ_ε^2. Let $S_j(d, \sigma_\varepsilon^2) = \sigma_\varepsilon^2 S_j'(d)$. Through the properties of diagonal and orthonormal matrices, the approximate log-likelihood function may be rewritten as

$$\widehat{\mathcal{L}}(d, \sigma_\varepsilon^2|\mathbf{x}) = N\log(\sigma_\varepsilon^2) + \log(S_{J+1}'(d)) + \sum_{j=1}^{J}\log(S_j'(d)) + \frac{1}{\sigma_\varepsilon^2}[\frac{\mathbf{v}_J^T\mathbf{v}_J}{S_{J+1}'(d)} + \sum_{j=1}^{J}\frac{\mathbf{w}_j^T\mathbf{w}_j}{S_j'(d)}]. \tag{4.35}$$

Differentiating Equation (4.35) with respect to σ_ε^2 and setting the result to zero, the MLE of σ_ε^2 is found to be

$$\widehat{\sigma}_\varepsilon^2 = \frac{1}{N}[\frac{\mathbf{v}_J^T\mathbf{v}_J}{S_{J+1}'(d)} + \sum_{j=1}^{J}\frac{\mathbf{w}_j^T\mathbf{w}_j}{S_j'(d)}].$$

Substituting σ_ε^2 into Equation (4.35) we obtain the reduced log-likelihood

$$\widehat{\mathcal{L}}(d|\mathbf{x}) = N\log(\sigma_\varepsilon^2) + \log(S_{J+1}'(d)) + \sum_{j=1}^{J}N_j\log(S_j'(d)).$$

The reduced log-likelihood is now a function of only the fractional differencing parameter $d \in (-1/2, 1/2)$.

This maximum likelihood (ML) procedure may be extended to allow for short memory (ARMA) components in the time series model – producing an ARFIMA(p,d,q) model given by

$$\Phi(B)(I-B)^d X_t = \Theta(B)\varepsilon_t,$$

where $\Phi(B)$ and $\Theta(B)$ are p and q degree polynomials in the backshift operator with autoregressive (AR) parameters $\mathbf{\Phi} = (\phi_1, \phi_2, \cdots, \phi_p)$ and moving average (MA) parameters $\mathbf{\Theta} = (\theta_1, \theta_2, \cdots, \theta_q)$, respectively. The approximate log-likelihood has been investigated by Jensen (1999a), who compared an approximate wavelet-based ML procedure to the approximate frequency-domain ML procedure in Fox and Taqqu (1986).

An alternative to the (approximate) ML estimators described above is to use an approximate wavelet ML (AWML) estimator. Following the arguments of McCoy and Walden (1996) and Johnstone and Silverman (1997), see also Jensen (1998, 2000), we assume that the asymptotic behavior (4.30) is satisfied, where σ^2 depends on other parameters of the model but do not vary with j.

The DWT provides a simple and effective method for approximately diagonalizing the variance/covariance of the original process. It follows that, ignoring wavelet coefficients $j > p-1-K$, the approximate wavelet likelihood function is given by

$$L_W(d, \sigma^2) = -\frac{1}{2} \sum_{j=0}^{p-1-K} [(2^j - 1)\log(\sigma^2 2^{-2jd}) + \sum_{k=0}^{2^j-1} \frac{\omega_{j,k}^2}{\sigma^2 2^{-2jd}}] \tag{4.36}$$

and the WML estimator is obtained by maximizing L_W. Since the relationship (4.30) is only valid for small j, we follow McCoy and Walden (1996) and Tse et al. (2002) and leave out the K largest scales in the likelihood function (4.36) to achieve robustness to the possible presence of short-run dynamics in the same sense as the semiparametric frequency domain estimators.

McCoy and Walden (1996) and Jensen (1999a) have both provided an approximate maximum likelihood (AML) estimator to the fractional difference parameter for long memory time series models.

4.2 Seasonal and/or Cyclical Long Memory (SCLM) Models

The potential value of a fractionally differenced model for a series, such as the monetary aggregates, is indicated by the sample autocorrelation function (ACF) of the first difference of the series. If a seasonally fractionally differenced model is appropriate, this ACF displays a hyperbolic decay at the seasonal lags, rather than the slow linear decay characteristic of the conventional seasonal differencing models, see in Figure 4.1 to have an intuitive idea. For this class of model, there are two types of long memory parameter estimation methods which are often used: a semiparametric one, based on the expression of

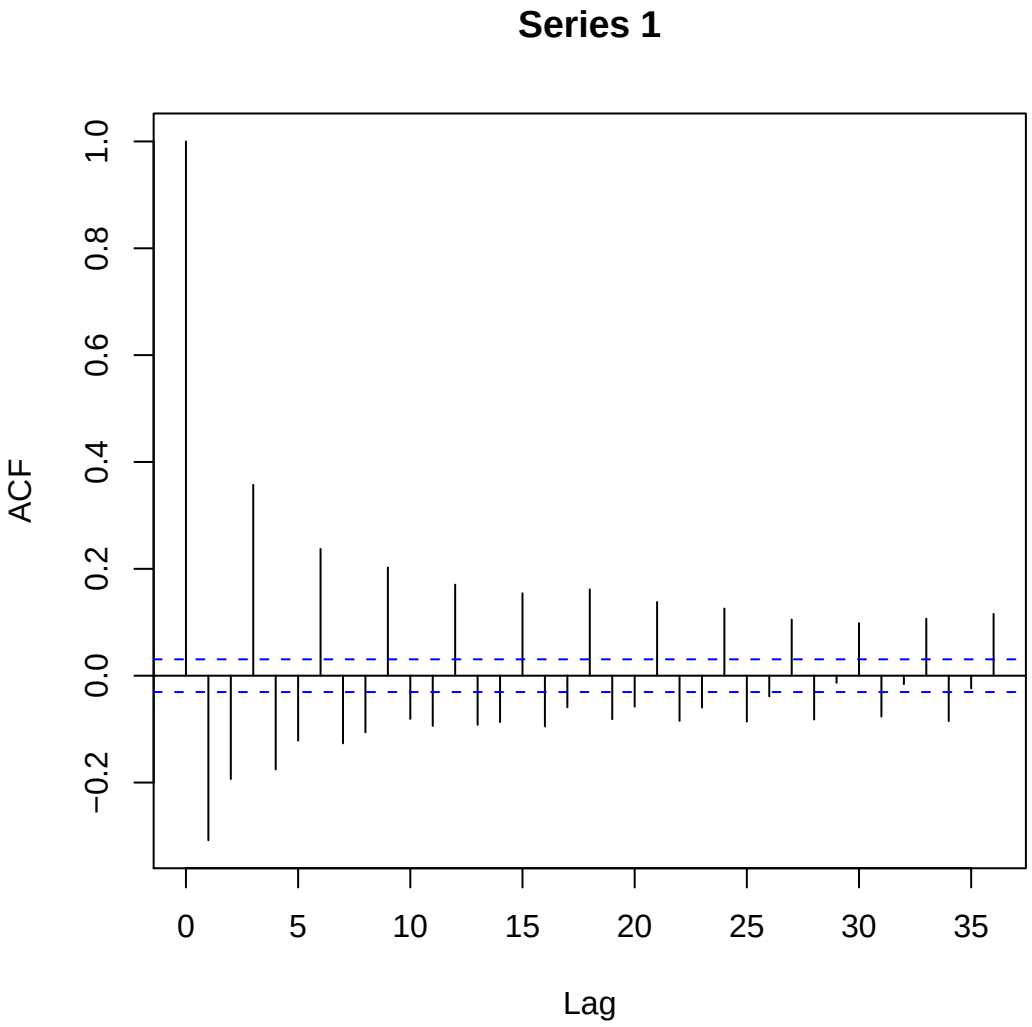


Figure 4.1: The ACF of a simulated seasonal long memory processes

the log-periodogram, and a pseudo maximum likelihood one, based on the Whittle likelihood.

One of the most difficult problems of the semiparametric method for the generalized long memory process with seasonalities is the choice of the trimming number l and the choice of the bandwidth m. The semiparametric method requires first the estimation of the Gegenbauer frequencies, and we use as estimates the values for which the peridogram is maximum. The semiparametric estimate has the great advantage to be easily computed, but it possesses nevertheless a slow convergence rate. Thus, this estimate must be carefully used in the case of small sample sizes, although statistical techniques, such as smoothing and tapering, can greatly improve the semiparametric estimate performances in such cases. Palma and Chan (2005) studied the asymptotic behavior of the estimated parameters of model (4.1) using pseudo-maximum likelihood method.

In practice, parameter estimation in statistical long memory models having spectral density with singularities is done in two steps (Gray et al., 1989; Chung, 1996a, b; Woodward et al., 1998). The first step consists in grid-search procedure to estimate the frequencies in which the spectral density is unbounded and in a second step, the memory parameter d is estimated by using a classical parametric method. Particularly, in the case of Gegenbauer process, Yajima (1996) proposed to estimate first the frequency of unbounded spectral density by maximizing the periodogram. The 2-steps method of Yajima (1996) and the simultaneous method have the great advantage to avoid a grid-search procedure, which is a very time-consuming method. Other authors considered a simultaneous global estimation of the whole of the parameters (Giraitis and Leipus, 1995; Ferrara, 2000). This method can be generalized to the case of a spectral density with several singularities (see Ferrara, 2000).

The common techniques for estimating the long-memory parameter for an ARFIMA model have been extended to the seasonal models, including log-periodogram and semi-parametric analysis (Arteche and Robinson, 2000). Ferrara (2000) used the Whittle's estimation of the maximum likelihood to estimate the parameters of k-factor Gegenauer processes. Whitcher (2004) applied the approximate maximum likelihood estimation to the case of long memory processes with seasonality, utilizing the DWPT under a particular basis function $\mathcal{B}$ to approximately diagonalize the variance/covariance matrix of the Gegenbauer process.

Recall the seasonal and/or cyclical long memory process (4.1) proposed by Robinson (1994):

$$(I-B)^{d_0}\sum_{i=1}^{k-1}(I-2\nu_i B+B^2)^{d_i}(I+B)^{d_k}X_t=\varepsilon_t.$$

We will consider in the following the estimation method for the particular cases of this general model.

4.2.1 Estimation for the k-factor Gegenbauer ARMA Processes

Hosking (1981) notes that taking the fractional power of a second-order polynomial makes it possible to describe long term structures of periodic shape. Thus, Gray et al. (1989) proposed the so-called Gegenbauer ARMA (GARMA) model, possessing a cyclical and persistent structure. Then Woodward et al. (1998) and Giratis and Leipus (1995) extended the GARMA model to the case in which the serie has a k cyclical persistent components, that is the k-factor Gegenbauer ARMA (k-factor GARMA) model. It is a particular case of the SCLM model.

Both time and frequency domain techniques have been established for the simulation of FI(d) process. The partitioning of the time-frequency plane by the DWT makes it a natural alternative to the discrete Fourier transform for decomposing long memory process. However, the DWT cannot be adapted to a general SDF of the Gegenbauer processes. In-

stead, we may make use of the DWPT, to produce the least correlated wavelet coefficients.

We consider first of all the stationary Gegenbauer process $(X_t)_t$ with a singularity on the interval $[0, \pi]$ in the spectrum. It contains only one persistent cyclical component. Let X_t be a stochastic process such that

$$(I - 2\nu B + B^2)^d X_t = \varepsilon_t \tag{4.37}$$

where ε_t is Gaussian white noise, then X_t is a Gegenbauer (or seasonal persistent) process, a simple example of the k-factor Gegenbauer ARMA (GARMA) process which permits the seasonalities and also takes into account the short memory terms.

Gray et al. (1989) showed that X_t is stationary and invertible for $|\nu| = 1$ and $-1/4 < d < 1/4$ or $|\nu| < 1$ and $-1/2 < d < 1/2$. When ν=1, Equation (4.37) becomes $(I - B)^{2d} X_t = \varepsilon_t$, which is an FI(d) process with parameter $2d$. Clearly, the definition of the Gegenbauer process also includes an FI(d) process.

Equation (4.37) may be rewritten as an infinite moving-average process via

$$X_t = \sum_{j=0}^{\infty} C_j(d, \nu)\varepsilon_{t-j},$$

where $C_j(d, \nu)$ are the Gegenbauer coefficients (Rainville, 1960) which can be calculated by the following recursion formula:

$$\begin{cases} C_0(d, \nu) &= 1 \\ C_1(d, \nu) &= 2d\nu \\ C_j(d, \nu) &= 2\nu(\dfrac{d-1}{j} + 1)C_{j-1}(d, \nu) - (2\dfrac{d-1}{j} + 1)C_{j-2}(d, \nu). \end{cases}$$

The SDF of X_t is given by

$$S_X(f) = \frac{\sigma_\varepsilon^2}{\{2|\cos(f) - \nu|^{2d}\}}, \quad for \ -\frac{\pi}{2} < f < \frac{\pi}{2}, \tag{4.38}$$

so that $S_X(f)$ becomes unbounded at frequency $f_G = \cos^{-1}(\nu)$, sometimes called the Gegenbauer frequency.

The autocovariance sequence (ACVS) of a Gegenabuer process may be expressed via the Fourier transform of its SDF; that is,

$$\gamma_\tau = \int_{-1/2}^{1/2} S_y(f)\cos(f\tau)df.$$

Gray et al. (1994) showed that the autocorrelation sequence of a Gegenabuer process is given by

$$\rho_\tau \sim \tau^{2d-1}\cos(f_G\tau), \quad as \ \tau \to \infty. \tag{4.39}$$

An obvious extension of the Gegenabuer process would be to allow multiple singularities to appear in the SDF of the process. Woodward et al. (1998) and Giraitis and Leipus (1995) considered the zero mean k-factor Gegenabuer process given by

$$\prod_{i=1}^{k}(I - 2\nu_i B + B^2)^{d_i} X_t = \varepsilon_t,$$

exhibiting k asymptotes located at the frequencies $f_i = \cos^{-1}(\nu_i)$ $(i = 1, \cdots, k)$, in its spectrum

$$S_X^{(k)} = \sigma_\varepsilon^2 \prod_{i=1}^{k} \{2|\cos(f) - \nu_i|\}^{-2d_i}, \ \ |f| < \frac{\pi}{2}. \tag{4.40}$$

Using this model allows one to incorporate several observed oscillations, such as an annual frequency and its harmonics. For the generalized long memory processes possessing the singularities at the non-zero frequencies in the spectrum, the DWT is not adaptable to decorrelate the general SDF.

Whitcher (2004) analyzed the feasibility of the OLS regression for the Gegenbauer processes. In fact, he established a linear relationship (on a log-log scale) between the DWPT variance and the long memory parameter, using the spectral density function of the Gegenbauer process. It is indeed a frequency-based semiparametric method. However, he showed that the wavelet-based OLS estimate of the fractional differencing parameter for Gegenbauer process exhibits reasonable bias and MSE characteristics in simulation studies. And if a non-adaptive orthonormal basis is used, the wavelet packet OLS estimator is heavily biased, since the basis mimic the frequency domain estimators of the SDF. Thus, he suggested that the wavelet-based OLS estimate should be restricted in its use in practice.

On the other side, Whitcher (2004) proposed the approximate maximum likelihood estimation (AMLE) for Gegenbauer process using the wavelet-based OLS estimate as the initial value. He utilized the DWPT under a particular choice of orthonormal basis to approximately diagonalize the covariance matrix of a Gegenbauer process. What's more, this maximum likelihood method can also be applied to the k-factor Gegenbauer ARMA processes possessing the short memory terms. Using the wavelet-based AMLE, he found that the bias and RMSE are largely improved.

- Ordinary Least Squares (OLS) Estimation of Gegenabuer Processes
  Porter-Hudak (1990) extended the non seasonal estimation technique developed in Geweke and Poter-Hudak (1983) to the fractionally differenced seasonal model and offered some preliminary sampling evidence as to its efficacy.

  And the observed linear relationship (on a log-log scale) between the wavelet variance $\sigma_x^2(\lambda_j)$ and scale λ_j provided the impetus to use OLS estimation for the fractional differencing parameter d of FI(d) process. Actually, there also exists a log-linear relationship for Gegenabuer process.

By applying the logarithmic transform to both sides of Equation (4.38), we get

$$\log S_X(f) = -2d \log 2|\cos(f) - \nu|. \tag{4.41}$$

This suggests a simple linear regression of $\log \widehat{S}_X(f)$ on $\log 2|\cos(2\pi f) - \nu|$ in order to estimate the fractional differencing parameter d, where $\log \widehat{S}_X(f)$ is an estimate of the true spectrum at each frequency (e.g. the periodogram or multitaper spectrum estimator). Arteche and Robinson (2000) suggested a simple modification to Equation (4.41), which consists of replacing the explicit parametric form of the SDF with just the frequency, yielding

$$\log S_X(f) \approx -2d \log 2|f - \cos^{-1}\nu|. \tag{4.42}$$

Thus, our variables for the OLS regression $\mathbf{y} = \mathbf{X}\beta + \mathbf{e}$ based on Equation (4.41) are

$$\mathbf{y} = \begin{pmatrix} \log \widehat{S}_X(f_1) \\ \log \widehat{S}_X(f_2) \\ \vdots \\ \log \widehat{S}_X(f_k) \end{pmatrix} \quad \mathbf{X} = \begin{pmatrix} 1 & \log 2|\cos(f_1) - \nu| \\ 1 & \log 2|\cos(f_2) - \nu| \\ \vdots & \vdots \\ 1 & \log 2|\cos(f_k) - \nu| \end{pmatrix} \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$$

and model errors $\mathbf{e} = [e_1, e_2, \cdots, e_J]^T$. The OLS estimator of β_1 is the second element of $\widehat{\beta} = (\mathbf{X^T X})^{-1}\mathbf{X^T y}$, thus using the transformation $\hat{d}_{OLS} = -\widehat{\beta}_1/2$ gives the estimated fractional differencing parameter. To utilize the regression based on Equation (4.42), simply construct the model matrix $\mathbf{X}$ using $\log 2|f - \cos^{-1}\nu|$ instead of $\log 2|\cos(f) - \nu|$. Additional modifications to this log-periodogram regression scheme may be found in Arteche and Robinson (2000). Robinson (1995) replaced the parametric form of the SDF with frequency in Equation (4.42) and found it to work quite well for long memory processes.

We now extend the results to formulate a wavelet packet variance estimator for the fractional difference parameter of Gegenabuer process. Since the variance of the wavelet coefficients is an estimation of the true spectrum, we obtain the following relationship:

$$\log \sigma^2(\lambda_{j,n}) = -2d \log 2|\cos(\mu_{j,n}) - \nu|, \tag{4.43}$$

where $\mu_{j,n}$ is the midpoint of the frequency interval $\lambda_{j,n}$. In strict terms, the (j,n)-th wavelet packet variance covers the entire interval of frequencies $\lambda_{j,n}$ but it suffices to represent this interval by its midpoint here. As in Equation (4.41), the slope from a simple linear regression of $\log \widetilde{\sigma}^2(\lambda_{j,n})$ on $\log 2|\cos(\mu_{j,n}) - \nu|$, appropriately normalized, provided an estimate of the fractional differencing parameter. Simplifying Equation (4.43) to just the frequencies, not the full SDF, yields

$$\log \sigma^2(\lambda_{j,n}) = -2d \log 2|\mu_{j,n} - \cos^{-1}\nu|. \tag{4.44}$$

For the wavelet packet variance, the variables for the OLS regression $\mathbf{y} = \mathbf{X}\beta + \mathbf{e}$ based on Equation (4.43) are

$$\mathbf{y} = \begin{pmatrix} \log \widehat{\sigma}_X(\lambda_{j,n}) \\ \log \widehat{\sigma}_X(\lambda_{j,n}) \\ \vdots \\ \log \widehat{\sigma}_X(\lambda_{j,n}) \end{pmatrix} \quad \mathbf{X} = \begin{pmatrix} 1 & \log 2|\cos(\mu_{j,n}) - \nu| \\ 1 & \log 2|\cos(\mu_{j,n}) - \nu| \\ \vdots & \vdots \\ 1 & \log 2|\cos(\mu_{j,n}) - \nu| \end{pmatrix} \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$$

and model errors $\mathbf{e} = [e_1, e_2, \cdots, e_J]^T$. As the log-periodogram estimator, the OLS estimator of β_1 is the second element of $\hat{\beta} = (\mathbf{X^T X})^{-1}\mathbf{X^T y}$ and using the transformation $\hat{d}_{OLS} = -\widehat{\beta}_1/2$ gives the estimated fractional difference parameter with $Var(\hat{d}_{OLS}) = \frac{1}{4}Var(\widehat{\beta}_1)$. To utilize the regression based on Equation (4.44), we simply construct the model matrix $\mathbf{X}$ using $\log 2|\mu_{j,n} - \cos^{-1}\nu|$ instead of $\log 2|\cos(\mu_{j,n}) - \nu|$.

Whereas the OLS estimate of the fractional differencing parameter for long memory processes has been shown to exhibit reasonable bias and MSE characteristics in simulation studies, Whitcher (2000) showed that the log-periodogram estimator using Equation (4.42) is heavily biased and the wavelet packet OLS estimate also performs poorly when using a non-adaptive orthonormal basis. This basis mimics frequency domain estimators of the SDF. When utilizing an adaptive orthonormal basis, the wavelet packet OLS estimate of the fractional differencing parameter outperforms both the log-periodogram and non-adaptive wavelet packet estimates. In practice, the OLS estimator $\hat{d}_{OLS}$ for Gegenabuer process should be restricted in its use to input into a maximum likelihood procedure and not be regarded as a visable estimator on its own.

- Approximate Maximum Likelihood Estimation of Gegenabuer Processes
  We have provided an approximate maximum likelihood estimator for the fractional differencing parameter of the long memory process. The DWT provides a simple and effective method for approximately diagonalizing the covariance matrix of the original process. We extend these results to the case of Gegenabuer process, where two parameters d and ν define the SDF. The key point is to utilize the DWPT under a particular choice of orthonormal basis $\mathcal{B}$ to approximately diagonalize the covariance matrix of a Gegenabuer process.

  Let $\mathbf{x}$ be a realization of a zero mean stationary Gegenbauer process with unknown parameter d, ν and $\sigma_\varepsilon^2 > 0$. Recall that the likelihood function of $\mathbf{x}$, under the assumption of multivariate Gaussianity, is given by

$$L(d, \nu, \sigma_\varepsilon^2|\mathbf{x}) = (2\pi)^{-N/2}|\Sigma_\mathbf{x}|^{-1/2}exp[-\frac{1}{2}\mathbf{x}^T\Sigma_\mathbf{x}^{-1}\mathbf{x}]. \tag{4.45}$$

  The MLEs of the parameters d, ν, and σ_ε^2 are those quantities that maximize Equation (4.45). We avoid the difficulties in computing the exact MLEs by using the

approximate decorrelation of the DWPT as applied to Gegenabuer process; that is,

$$\Sigma_{\mathbf{x}} \approx \widehat{\Sigma}_{\mathbf{x}} = \mathcal{W}_{\mathcal{B}}{}^{T}\Omega_{N}\mathcal{W}_{\mathcal{B}}$$

where $\mathcal{W}_{\mathcal{B}}$ is an $N \times N$ orthonormal matrix defining the DWPT through the basis $\mathcal{B}$ and Ω_N is a diagonal matrix containing the band-pass variances associated with $(j,n) \in \mathcal{B}$. The approximate likelihood function is now

$$\widehat{L}(d,\nu,\sigma_{\varepsilon}^{2}|\mathbf{x}) = (2\pi)^{-N/2}|\widehat{\Sigma}_{\mathbf{x}}|^{-1/2}exp[-\frac{1}{2}\mathbf{x}^{T}\widehat{\Sigma}_{\mathbf{x}}^{-1}\mathbf{x}].$$

Hence we want to find values of d, ν and σ_{ε}^{2} that minimize the log-likelihood function

$$\widetilde{\mathcal{L}}(d,\nu,\sigma_{\varepsilon}^{2}|\mathbf{x}) = -2\log(\widehat{L}(d,\nu,\sigma_{\varepsilon}^{2}|\mathbf{x})) - N\log(2\pi) = \log(|\widehat{\Sigma}_{\mathbf{x}}|) + \mathbf{x}^{T}\widehat{\Sigma}_{\mathbf{x}}^{-1}\mathbf{x}.$$

We know that the variance for DWPT coefficients associated with the frequency interval $\lambda_{j,n}$ is given by $S_{j,n}$. We note that $S_{j,n}$ depends on three parameters related to the Gegenabuer process: the fractional difference parameter d, the Gegenbauer frequency $f_G = \cos^{-1}(\nu)$, and the variance σ_{ε}^{2}. Let $S_{j,n}(d,\nu,\sigma_{\varepsilon}^{2}) = \sigma_{\varepsilon}^{2}S'_{j,n}(d,\nu)$. Since properties of diagonal and orthonormal matrices, the approximate log-likelihood function may be rewritten as

$$\widetilde{\mathcal{L}}(d,\nu,\sigma_{\varepsilon}^{2}|\mathbf{x}) = N\log(\sigma_{\varepsilon}^{2}) + \sum_{(j,n)\in\mathcal{B}} N_j \log(S'_{j,n}(d,\nu)) + \frac{1}{\sigma_{\varepsilon}^{2}}[\sum_{(j,n)\in\mathcal{B}} \frac{\mathbf{w}_{j,n}^{T}\mathbf{w}_{j,n}}{S'_{j,n}(d,\nu)}]. \tag{4.46}$$

Differencing Equation (4.46) with respect to σ_{ε}^{2} and setting the result to zero, the MLE of σ_{ε}^{2} is found to be

$$\widehat{\sigma}_{\varepsilon}^{2} = \frac{1}{N}[\sum_{(j,n)\in\mathcal{B}} \frac{\mathbf{w}_{j,n}^{T}\mathbf{w}_{j,n}}{S'_{j,n}(d,\nu)}].$$

Substituting $\widehat{\sigma}_{\varepsilon}^{2}$ into Equation (4.46), we obtain the reduced log-likelihood function as follows:

$$\widehat{\mathcal{L}}(d,\nu|\mathbf{x}) = N\log(\widehat{\sigma}_{\varepsilon}^{2}) + \sum_{(j,n)\in\mathcal{B}} N_j \log(S'_{j,n}(d,\nu)).$$

This estimation procedure differs from the frequency-based semiparametric estimator of Arteche and Robinson (2000) by simultaneously determining MLEs for both the fractional difference parameter and Gegenbauer frequency.

The estimation procedure outlined here has assumed only one singularity in the spectrum of X_t. It is common to observe a fundamental frequency, say the annual cycle, and several harmonics, such as cycles of two per year, and so on. These may be included in the ML procedure by using the spectrum of a k-factor Gegenabuer process. The likelihood function would then be a function of $\mathbf{d} = (d_1, d_2, \cdots, d_k)$, $\mathbf{\Phi} = (\phi_1, \phi_2, \cdots, \phi_k)$ and σ_{ε}^{2}. Long memory may also be incorporated into the

modeling by allowing one of the Gegenabuer frequencies to be zero. The fractional difference parameter associated with this zero frequency would be constrained via $|d| \leq 1/4$, but the relation to a fractional differencing parameter of an FI(d) is $2d$. Finally, short memory may also be included by adding AR or MA terms to the spectrum of the model. Thus we allow ML estimation for parameters to cover a wide variety of time series models involving long memory with seasonality.

The spectral density of the k-factor GARMA process is as follows:

$$S_X(f) = \frac{\sigma_\varepsilon^2}{2\pi} \frac{|\Theta(e^{if})|^2}{|\Phi(e^{if})|^2} \prod_{i=1}^{k} |4\sin(\frac{f+f_i}{2})\sin(\frac{f-f_i}{2})|^{-2d_i}, \quad for\ f \in [0, 2\pi), \tag{4.47}$$

where $f_i = \cos^{-1}(\nu_i)$ $(i = 1, \cdots, k)$ are the k Gegenbauer frequencies. We have the approximation of the spectral density as follows:

$$S_X(f) \sim C|f - f_i|^{-2d_i^*}, \quad when\ f \to f_i, \tag{4.48}$$

where C is a strictly positive finite constant and

$$d^* = \begin{cases} 2d_i, & if\ f_i \in \{0, \pi\}, \\ d_i, & if\ 0 < f_i < \pi. \end{cases}$$

Arteche (1998) carried out the generalized Robinson method with the trimming technique to estimate $d_i (i = 1, \cdots, k)$ and proved the asymptotic behavior of the estimates.

For the k-factor GARMA model, the estimation method proposed by Chung (1996a, b) is based on the minimization of the conditional sum of squared residuals (CSS). However, parameter estimation of this class of process is delicate. Chung (1996a, b) showed that the estimator of ν obtained by CSS minimization converges at a greater speed than the other parameters. This rules out the use of gradient-based methods on the whole set of parameters. It is therefore advisable to use an alternative method based on an incremental search (or grid-search). This method is very slow if the grid-search corresponds to $[-1, 1]$. It is then more effective to restrict the search interval to a neighborhood of frequencies relative to the strongest values of the periodogram. Chung demonstrated that the distribution of the estimators of d, as well as that of a possible ARMA structure obtained by minimization of the CSS function, is asymptotically normal. Parameter significance can be tested by a significant Student test.

Recall that the Whittle estimator of $\theta = (d, \nu, \sigma_\varepsilon^2)$ is obtained by minimizing the following approximation of the log-likelihood function (see Beran 1994):

$$L_W(X, \theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} (\log(f_X(\lambda, \theta)) + \frac{I_T(\lambda)}{f_X(\lambda, \theta)}) d\lambda, \tag{4.49}$$

where $f_X(\lambda, \theta)$ is the spectral density of the k-factor Gegenbauer process $(X_t)_t$ generating the data, and $I_T(\lambda)$ is the periodogram defined by

$$I_T(\lambda) = \frac{1}{2\pi T} |\sum_{t=1}^{T} e^{i\lambda t}(X_t - \bar{X}_T)|^2,$$

where $\bar{X}_T$ is the empirical mean of the process, equal to zero in our case.

Under some classical conditions inherent to the Whittle estimation procedure, Giraitis and Leipus (1995) proved the strong consistency of the Whittle estimator $\hat{\theta}_T$, by using a result of Hannan (1973). The limiting distribution of the Whittle estimator of Gegenbauer process was proven by Diongue and Guégan (2004). The asymptotic Gaussian distribution of the Whittle estimator has been proved under some non-restrictive conditions, by extending the proof of Yajima (1985) to the vectorial case.

It is worthwhile to note that the convergence rate of this pseudo maximum likelihood estimate is greater than the convergence rate of the semiparametric estimate. While the convergence rate of the conditional sum of squares (CSS) estimate of the Gegenbauer frequency is $O(T^{-1})$. Whitcher (2004) proposed a method to estimate the parameters of the Gegenbauer processes using semiparametric method by wavelet transformation. He took the logarithmic of the SDF of Gegenbauer process, yielding

$$\log S_X(f) = -2d \log |\cos f - \nu|.$$

Thus, we obtained a simple linear regression of $\log \hat{S}_X(f)$ to estimate the long memory parameter d, where $\hat{S}_X(f)$ is an estimate of the true spectrum at each frequency.

4.2.2 Estimation for the Models with Fixed Seasonal Periodicity

Consider the simple model

$$(I - B^s)^d X_t = \varepsilon_t, \tag{4.50}$$

where d is the fractionally differenced component and lies inside the interval $(-1/2, 1/2)$, ε_t is a white noise process, and s is the seasonal periodicity (e.g. $s = 12$ for monthly series). The model in (4.50) is a direct analogue of the simple fractional differenced model (4.23). The generalization of (4.50) to an autoregressive moving average (ARMA) model with a fractionally differenced seasonal component is

$$\Phi(B)(I - B^s)^d X_t = \Theta(B)\varepsilon_t, \tag{4.51}$$

where $\Phi(B)$ and $\Theta(B)$ are autoregressive and moving average polynomials (each conceivably including seasonal components).

Porter-Hudak (1990) extended the results of Geweke and Porter-Hudak (1983) to the model (4.51), considering the spectral density of the model (4.50),

$$S_X(f) = \sigma^2 / 2\pi (2(1 - \cos(sf)))^{-d}. \tag{4.52}$$

The natural extension of the non-seasonal technique given in Geweke and Porter-Hudak (1983) is a regression of the log periodogram, $\log I(\pi j/T)$, around the seasonal harmonics, $\pi sj/T (j = 1, 2, \cdots, g(T))$, for some choice of $g(T)$ such that the limit of $g(T)/T$ goes to 0 as $T \to \infty$,

$$\log I(\pi j/T) = \phi_0 - d \log(2(1 - \cos(\pi js/T))), \;\; j = 1, 2, \cdots, g(T). \tag{4.53}$$

Consider a more general seasonal process. Porter-Hudak (1990) introduced the following model, taking into account the presence of an infinite cycle as well as a given seasonality with a certain persistence:

$$\Phi(B)(I-B)^{d_1}(I-B^s)^{d_2}X_t = \Theta(B)\varepsilon_t, \tag{4.54}$$

whose spectral density is given by, for $-\pi \leq f \leq \pi$:

$$S_X(f) = \frac{\sigma^2}{2\pi}(2\sin(fs/2))^{-2d_2}(2\sin(f/2))^{-2d_1}. \tag{4.55}$$

The parameter estimation is generally done with the Geweke and Porter-Hudak (1983) method, or GPH method. Two fractionally differenced components d_1 and d_2 could be estimated in a multivariate regression setting, by considering the following regression:

$$\log I(\pi j/T) = \phi_0 - d_1 \log(2(1-\cos(\pi j/T))) - d_2 \log(2(1-\cos(\pi js/T))),\ j = 1, 2, \cdots, g(T). \tag{4.56}$$

Unfortunately, this latter expression is asymptotically equivalent to $\log(\pi j/T) = \phi_0 - d_1 \log(\pi j/T) - d_2 \log(\pi js/T)$ or $\log I(\pi j/T) = \phi - (d_1 + d_2)\log(\pi j/T)$. Hence both d_1 and d_2 can not be identified (see Porter-Hudak, 1983).

4.3 Seasonal and/or Cyclical Asymmetric Long Memory (SCALM) Models

By adding the asymmetry, Arteche and Robinson (2000) introduced the Seasonal and/or Cyclical Asymmetric Long Memory model, or SCALM. They consider a semiparametric approach using spectral density defined in the following way:

$$f(\lambda) \sim C_1|\lambda - \omega| \quad as\ \ \lambda \to \omega^+, \tag{4.57}$$

$$f(\lambda) \sim C_2|\lambda + \omega| \quad as\ \ \lambda \to \omega^-, \tag{4.58}$$

where $\omega \in (0, \pi]$ and for $i = 1, 2$,

$$0 < C_i < \infty,\ |d_i| < \frac{1}{2}, \tag{4.59}$$

permitting

$$d_1 \neq d_2,\ and/or\ C_1 \neq C_2. \tag{4.60}$$

Arteche and Robinson (2000) proposed an estimate of d_1 and d_2 based on a trimming approach of the periodogram. A complementary approach for parameter estimation is considered by Olhede, McCoy and Stephens (2004).

Chapter 5

Estimation and Forecast for Non-stationary Long Memory Processes

Stationarity has always played an important role in time series analysis. One reason is that for stationary processes there exists a rich and elegant theory which allows for a detailed investigation of the different methods used in statistical inference. As a consequence, practitioners often try to use stationary methods even when the data clearly show a non-stationary behavior, e.g. by taking differences or by looking at different segments separately.

However, in some situations, the assumption that real world processes exhibit a constant long memory structure may not be reasonable. Time-varying long memory characteristics have been hypothesized or observed in telecommunications networks, physiological signals, seismic measurements, etc. To better capture the non-stationary behavior associated with market collapses, political upheavals and new announcements, we need to study the non-stationary models with the time-varying parameters.

It is difficult to develop a general non-stationary theory. On one hand, an asymptotic theory is needed, since for example an investigation of a maximum likelihood estimate for fixed sample size is much too complicated and will not lead to any satisfactory results. On the other hand, a classical asymptotic theory with the assumption that more and more observations of the future become available does not make sense since future observations of a general non-stationary process do not necessarily contain any information of the structure at present.

It is well-known that the Fourier transform is well-equipped with its localized frequency basis functions and spectral representation to deal with stationary time series, but its use in non-stationary setting is not recommended. Whereas wavelet analysis is more appropriate to analyze the time-varying behavior of non-stationary time series because it is well localized both in time and scale. Compared with the study on wavelet applications of stationary long memory processes, the wavelet-based work on non-stationary processes is not so much. Here, in order to provide a consistent and robust procedure to estimate the long memory parameters of non-stationary processes, we focus on wavelet techniques.

Summarizing the time-varying nature through a constant long memory parameter, also known as the Hurst coefficient H, yields a stationary model that does not capture its non-stationary behavior. Gonçalvès and Abry (1997) estimated a local scaling exponent for a continuous-time, multifractal, Brownian motion process characterized by a time-varying Hurst coefficient $H(t)$. Their method relies on the multiple window scalogram (squared magnitude of the continuous wavelet transform) to estimate $H(t)$. This involves constructing non-standard wavelets to compute the scalograms, which may hinder practical implementation.

As far as we know, the existing work concerning the non-stationary processes is mostly concentrated on the fractional integrated process $(I - B)^d X_t = \varepsilon_t$ with a constant differencing parameter $d \in (\frac{1}{2}, 1)$ and the fractional model $(I - B)^{d(t)} X_t = \varepsilon_t$ with a time-varying parameter function $d(t) \in (-\frac{1}{2}, \frac{1}{2})$.

For the first case, many authors adopted the estimation method based on the Whittle method, see for example Abadir et al. (2007), Beran and Terrin (1996), Moulines et al. (1998), Phillips and Shimotsu (2000, 2004, 2006) and Velasco and Robinson (2000). For the second case, the possibility that the long memory parameter d is not constant over time is an interesting generalization of the usual FI(d) process. Veitch and Abry (1999) developed a testing procedure for the time constancy of d, while Whitcher and Jensen (2000) proposed an OLS estimator for a non-stationary FI(d) process, and Jensen and Whitcher (2000) applied the OLS-based estimator to a year of high-frequency foreign exchange rates. Parameter estimation for a non-stationary long memory time series model, through OLS or maximum likelihood, is in its infancy and should benefit greatly from wavelet-based methods.

In the literature, in order to estimate the fractional differencing parameter of the processes for which the spectrum has the singularity only at zero, the authors perform the orthonormal discrete wavelet transform (DWT) to decorrelate the original time series (Jensen, 1999a, b; Cavanaugh et al., 2002). The band-pass structure of the DWT partitions the spectrum finer and finer as the frequency tends to 0, where the spectrum is unbounded. It is performed through a succession of low-pass and high-pass filtering operations.

For the long memory processes with seasonal terms, the spectral density function can explode at any frequency between 0 and π. We apply the discrete wavelet packet transform (DWPT) instead of DWT which permits to approximately decorrelate the spectrum of the process. To realize this approximate decorrelation, we resort to the minimum-bandwidth discrete-time (MBDT) wavelets with length L (denoted by MB(L)), introduced by Morris and Peravali (1999). It permits to approximately decorrelate the band-pass filter and to choose the adaptive orthonormal basis (Whitcher, 2004).

Since there exists an OLS regression (4.43) to obtain the long memory parameter of the Gegenbauer process, we will try to utilize this method in the estimation of non-stationary

long memory processes with seasonalities, see Chapter 5 for details. And the approximate likelihood estimation method for non-stationary long memory processes with seasonalities is still under consideration.

5.1 Fractional Integrated Processes with a Constant Long Memory Parameter

Moulines et al. (2008) considered a time series with memory parameter $d \in \mathbb{R}$. This time series is either stationary or can be made stationary after differencing a finite number of times. They proposed a wavelet-based semiparametric pseudo-likelihood maximum method estimator – the local Whittle wavelet estimator of the memory parameter d. They also showed that the estimator is consistent and rate optimal if the process is linear, and is asymptotically normal if the process is Gaussian.

In fact, the study of the estimation of long memory parameters in non-stationary long memory processes is mostly focused on the fractional integrated process proposed by Granger and Joyeux (1980) and Hosking (1981) to take into account the existence of non-stationarity jointly with presence of persistence.

Consider the model

$$(I - B)^d X_t = \varepsilon_t, \ \ |d| \geq \frac{1}{2}. \tag{5.1}$$

Such fractional integrated process has been studied by Beran and Terrin (1996), Velasco and Robinson (2000), Shimotsu and Phillips (2000, 2004, 2006), Abadir et al. (2007) and Moulines et al. (2008), for instance. In these work, the authors applied Geweke-Porter-Hudak (GPH) method, local Whittle method, exact local Whittle method, fully extended local Whittle method, Whittle pseudo maximum likelihood method, wavelet-based local Whittle method for estimating the parameter.

Considering the fractional integrated process (5.1), Phillips and Shimotsu (2001) found that for the local Whittle (LW) estimator of the non-stationary fractional integrated processes ($|d| \geq \frac{1}{2}$), the asymptotic theory is discontinuous at $d = \frac{3}{4}$ and $d = 1$. Thus it is awkward to use the LW estimator because of non-normal limit theory and the estimator is inconsistent when $d > 1$. And the LW estimator is not a good general purpose estimator when the value of d may take on values in the non-stationary zone beyond $\frac{3}{4}$. Shimotsu and Phillips (2002) proposed the exact local Whittle method for the estimation of the fractional integration, which is applicable to the stationary and non-stationary processes. And the estimator is shown to be consistent and to follow the $N(0, \frac{1}{4})$ limit distribution for all values of d.

When $d \in (-\frac{3}{2}, \infty)$, Abadir et al. (2007) investigated the properties of fully extended local Whittle (FELW) estimator, which is applicable not only for the traditional cases but

also for nonlinear and non-Gaussian processes. They showed that the estimator is consistent and has the good asymptotic behavior.

Whittle pseudo-maximum likelihood estimates of parameters for stationary time series have been found to be consistent and asymptotically normal in the presence of long range dependence. Generalizing the definition of the long memory parameter d, Velasco and Robinson (2000) extended these results to include possibly non-stationary $(0.5 \leq d < 1)$ and anti-persistent $(-0.5 < d < 0)$ observations. Using adequate data tapers , we can apply this estimation technique to any degree of non-stationarity $d \geq 0.5$ without prior knowledge of the memory of the series.

5.2 Locally Stationary ARFIMA Processes

For the non-stationary long memory processes, some authors studied the processes with time-varying long memory parameters instead of a constant long memory parameter. For instance, a locally stationary ARFIMA model is studied in Jensen (1999a) and Whitcher and Jensen (2000). They introduced the operator $(I-B)^{d(t)}$. And Cavanaugh et al. (2002) investigated the self-similar processes, for example, the time-varying fractional Brownian motion and the fractional Gaussian noise with the time-varying parameter. The authors developed the estimation procedures based on wavelets techniques, succeeding in capturing the local changes in the series.

Existing frequency-domain estimators of the long memory parameter depend on the Fourier transforms localized in frequency. These estimators are incapable of addressing any time-varying long memory behavior (see Geweke and Porter-Hudak, 1983; Fox and Taqqu, 1986) for two of the most popular frequency domain, long memory estimators). By definition, the statistical properties of a non-stationary process are a function of time and scale. Since the wavelet is localized in time, it is feasible to focus on a period where the statistical properties of the non-stationary process are relatively stable and not affected by observations with differing statistical properties. Hence, whereas Fourier analysis is for the study of stationary processes, wavelet analysis is more suitable for the study of non-stationary processes.

Whitcher and Jensen (2000) introduced a stochastic process defined $X_{t,T}$ given by

$$X_{t,T} = (I - B)^{-d(t)} \frac{\Theta(B)}{\Phi(B)} \varepsilon_t \equiv \sum_{j=0}^{\infty} \frac{\Gamma[j + d(t)]}{\Gamma(j+1)\Gamma[d(t)]} \Psi_j \varepsilon_{t-j}, \tag{5.2}$$

where $d(t)(\in (-\frac{1}{2}, \frac{1}{2}))$ is the time-varying fractional differencing parameter and ε_t is a sequence of mean zero normal (Gaussian) random variables with variance σ_ε^2. Here, B denotes the lag (backshift) operator, that is $X_{t-j,T} = B^j X_{t,T}$, and $\Gamma(\cdot)$ is the gamma function. The functions $\Theta(B)$ and $\Phi(B)$ are respectively p and q order polynomials in the lag operator B, each with roots outside the unit circle, and Ψ_j is the solution to

$\Psi_j z^j = \dfrac{\Theta(z)}{\Phi(z)}$. Called a locally stationary ARFIMA model, this long memory time series model is a member of the non-stationary class of processes known as locally stationary processes (Dalhaus, 1996).

Actually, we can rewrite the model (5.2) in the following form:

$$\Phi(B)(I-B)^{d(t)}X_{t,T} = \Theta(B)\varepsilon_t. \tag{5.3}$$

This process can be regarded as an extension of the ARFIMA model allowing the long memory parameter to evolve over time.

The time-varying spectral density function for $X_{t,T}$ is given by $S(u,\omega) \sim \omega^{-2d(u)}$, where $u = t/T$. If $d(u) > 0$, $S(u,\omega)$ is smooth for frequencies close to zero, but is unbounded when $\omega = 0$. In other words, the energy of $X_{t,T}$ is concentrated over those frequencies associated with long term cycles. If $d(u) < 0$, then $S(u,0) = 0$ and $X_{t,T}$ is a locally stationary series that is anti-persistent. As a result of the time-varying long memory parameter, $X_{t,T}$ will be smoother with less variation in its amplitude during time periods where $d(u) > 0$. And $X_{t,T}$ will have large fluctuations in its value when $d(u) < 0$.

The short memory parameter found in $\Theta(B)$ and $\Psi(B)$ could be modeled as functions of t. Since these parameters only affect the short-run dynamics of the process and our main interest is to study the estimation of the long memory parameters, thus we can set the short memory parameters to be zero.

Whitcher and Jensen (2000) extended Jensen (1999a)'s wavelet-based ordinary least squares (OLS) estimator of long memory d to $d(t)$, by using the cone of influence to determine the location in time of the long memory estimate. They introduced an estimation procedure for the fractional differencing parameter function of a particular model, locally stationary long memory processes. Instead of arbitrarily partitioning the data, they allowed the support of the central portion of the wavelet filter to determine a scale-dependent window for computing the local wavelet variance. The estimator is calculated via ordinary least squares regression applied to the local wavelet variances. The wavelet-based estimator of the local long-memory parameter is demonstrated using vertical ocean shear data.

Besides, Cavanaugh et al. (2002) considered the model (5.3) as a locally self-similar process with the scaling function $H(t) = d(t) + \dfrac{1}{2}$, since many phenomena exhibit self-similar patterns that change as the phenomenon itself evolves. They proved that the model (5.3) is a locally self-similar process, then by applying the discrete wavelet transform and by partitioning the time interval, they built the approximate log-linear relationship between the wavelet coefficients and the scale, and they carried out the OLS regression on each partitioned subinterval, thus they obtained the local estimates of the scaling parameter function. After smoothing, they got the shape of the parameter function. And they also proved the consistency of the estimator. Then several practical applications are presented, for example in the data of the vertical ocean shear measurements, yearly water levels for Nile river and Ethernet network traffic measurements, which illustrated that

their wavelet-based method can quantify time-dependent self-similarity patterns that arise in actual temporal and spatial series.

5.3 Locally Stationary k-factor Gegenbauer Processes

Long memory processes have been extensively studied over the past decades. When we deal with the financial and economic data, seasonality and time-varying long-range dependence can often be observed and thus some kind of non-stationarity can exist inside financial data sets. To take into account this kind of phenomena, we propose a new class of stochastic process: the locally stationary k-factor Gegenbauer process. We describe a procedure of estimating consistently the time-varying parameters by applying the discrete wavelet packet transform (DWPT). The robustness of the algorithm is investigated through simulation study. We propose also the forecast method for this new non-stationary model.

Working with the existence of non-stationarity does not mean that we observe explosions. Due to the existence of seasonalities inside data sets, the k-factor GARMA model has been extensively applied in economics and finance, which was justified by Diebold and Rudebush (1989), Sowell (1992), Gil-Alana and Robinson (2001), Carlin and Dempster (1989), Porter Hudak (1990), Ray (1993), Franses and Ooms (1997), Arteche and Robinson (2000), Arteche (2003), Darné et al (2004), Ferrara and Guégan (2001a, b) and Gil-Alana and Hualde (2008), to name just a few.

Most of the applications within this framework assume that the data sets are stationary. But in practice, series cannot always be made stationary even by transformation or sometimes it makes no sense to render the data sets stationary. Generally, that the underlying process is mean-reverting can be useful in practice. Some examples can be found in ecology for instance, Whitcher and Jensen (2000) and Cavanaugh et al. (2002).

In this part, we extend the k-factor Gegenbauer model assuming that the parameters d_i evolve with time. Thus, we introduce a new model which is locally stationary taking into account both the presence of persistence and the existence of seasonalities, and we provide an estimation procedure adapted to this new class of models. We consider the following model:

$$\prod_{i=1}^{k}(I - 2\cos\lambda_i B + B^2)^{d_i(t)} y_t = \varepsilon_t, \tag{5.4}$$

where ε_t is a Gaussian white noise. In this model, the long memory parameter is time-varying, thus the model is non-stationary, which can be regarded as piecewise stationary process. In the spectrum of this model, we can also observe the explosions at non-zero frequency, which indicates the seasonality.

We specify now why the locally stationary k-factor Gegenbauer model is well-defined and the conditions which ensure the local stationarity. First of all, we remark that the

operator $\prod_{i=1}^{k}(I - 2\cos\lambda_i B + B^2)^{d_i(t)}$ permits to define the fixed frequencies λ_i of the spectrum characterized by the time-varying long memory parameters $d_i(t)$ $(1 \leq i \leq k)$. Then, we can expand the previous operator in the following way:

$$\prod_{i=1}^{k}(I - 2\cos\lambda_i B + B^2)^{d_i(t)} = \sum_{n=0}^{\infty}\pi_n(t)B^n, \tag{5.5}$$

and the coefficients $\pi_n(t)$ verify:

$$\pi_n(t) = \sum_{\substack{0 \leq j_1, \cdots, j_k \leq n \\ j_1 + \cdots + j_k = n}} c_{j_1}^{(-d_1(t))}(\cos\lambda_1)\cdots c_{j_k}^{(-d_k(t))}(\cos\lambda_k) \tag{5.6}$$

in which $c_k^{(d(t))}(x)$ are orthogonal Gegenbauer (or ultraspherical) polynomials defined on $[-1, 1]$. Thus, we get the following proposition which ensures the existence and locally stationarity conditions for the locally stationary k-factor Gegenbauer process.

Proposition 5.3.1. *Let $d_j(t)(j = 1, \cdots, k)$ be regular and nonzero functions satisfying the condition*

$$|d_j(t)| < \begin{cases} \frac{1}{2}, & if \quad 0 < \lambda_j < \pi \\ \frac{1}{4}, & if \quad \lambda_j = 0 \text{ or } \pi. \end{cases} \tag{5.7}$$

Then, there exists a unique solution for the locally stationary k-factor Gegenbauer model which has the following representation

$$y(t) = \sum_{n=0}^{\infty}\psi_n(t)\varepsilon(t-n). \tag{5.8}$$

The coefficients $\psi_n(t)$ verify:

$$\begin{aligned}\psi_n(t) &= 2\sum_{k:0<\lambda_k<\pi} D(k,t)\frac{\Gamma(n + d_k(t))}{\Gamma(n+1)\Gamma(d_k(t))}\cos(\lambda_k n + \nu_k) \\ &+ \sum_{k:\lambda_k = 0 \text{ or } \pi} D(k,t)\frac{\Gamma(n + 2d_k(t))}{\Gamma(n+1)\Gamma(2d_k(t))}\cos(\lambda_k n) + O(n^{-2+\max\{d_1^*(t),\cdots,d_k^*(t)\}})\end{aligned} \tag{5.9}$$

for any given time t, as $n \to \infty$, where

$$d_k^*(t) = \begin{cases} d_k(t), & if\ 0 < \lambda_k < \pi \\ 2d_k(t), & if\ \lambda_k = 0 \text{ or } \pi, \end{cases}$$

$$\nu_k(t) = \lambda_k \sum_{j=1}^{k} d_j(t) - \pi\sum_{j=1}^{k-1} d_j(t) - d_k(t)\frac{\pi}{2},$$

and

$$D(k,t) = \begin{cases} |2\sin\lambda_k|^{-d_k(t)} \prod_{j\neq k} |2(\cos\lambda_k - \cos\lambda_j)|^{-d_j(t)}, & if\ 0 < \lambda_k < \pi. \\ \prod_{j\neq k} |2(\cos\lambda_k - \cos\lambda_j)|^{-d_j(t)}, & if\ \lambda_k = 0\ or\ \pi. \end{cases}$$

The prove of the proposition can be referred to in the Appendix.

This proposition ensures that the locally stationary k-factor Gegenbauer process is well-defined since every L^2 functions can be approximated by regular functions. Thus, for the locally stationary k-factor Gegenbauer process, the parameters $d_i(t)$ evolve with time and, at each time $d_i(t)$ are constant. Therefore locally, the locally stationary k-factor Gegenbauer model corresponds to the stationary k-factor Gegenbauer process.

5.3.1 Procedure for Estimating $d_i(t)$

In the stationary case, estimation methods for long memory models with seasonalities have been developed using semiparametric methods (Robinson, 1995; Chung, 1996a, b; Arteche and Robinson, 2000; Diongue et al., 2004).

In this part, we develop a procedure based on wavelet method in order to estimate the time-varying long memory parameter $d_i(t)$. We focus on wavelet method to investigate properties of this new model using the discrete wavelet packet transform (DWPT) which permits to approximately decorrelate the spectrum of the process. We first establish an approximate log-linear relationship between the time-varying wavelet variance of the DWPT coefficients and the time-varying long memory parameter $d_i(t)$. Finally we apply locally the OLS regression method to obtain local estimates of the time-varying parameters. Our method may be regarded as an extension of the log-linear regression techniques proposed by Whitcher (2004). It can also be considered as an extension of the estimation technique of locally self-similar parameters proposed by Cavanaugh et al. (2002).

5.3.2 Estimation Procedure

First, we assume that the sample size is dyadic ($N = 2^J$), otherwise we repeat the last data value several times to achieve such a sample size.

In the first step, we are restricted to a locally stationary 1-factor Gegenbauer model and we assume that we observe y_t ($t = 1, \cdots, N$), such that:

$$(I - 2\nu B + B^2)^{d(t)} y_t = \varepsilon_t, \tag{5.10}$$

$(\varepsilon_t)_t$ being a Gaussian white noise. We assume that the time-varying fractional differencing parameter is such that $|d(t)| < 1/2$, B is the backshift operator, $y_{t-j} = B^j y_t$. In order to estimate $d(t)$, providing an asymptotic theory, we need to make N tend to infinity. To

avoid instability of $d(t)$, we suppose that we observe $d(t)$ on a finer grid (making $d(t)$ rescaled on $[0,1]$), that we observe $(y_{t,N})$ such that:

$$(I - 2\nu B + B^2)^{d(t/N)} y_{t,N} = \varepsilon_t. \tag{5.11}$$

Letting N tends to infinity means that we have in the sample $y_{1,N}, \cdots, y_{N,N}$ more and more observations for each value of $d(t)$.

Now, we are going to characterize this local stationary process through its spectral density: this tool is superior in presence of seasonalities, to the autocovariance function.

The stochastic process defined in (5.10) is a Gegenbauer process that is locally stationary in the sense of Dalhaus (1996a), with realizations of length N. Its spectral density is such that:

$$f_N(\lambda) = \frac{\sigma^2_{\varepsilon_t}}{2\pi} \frac{1}{(2|cos\lambda - \nu|)^{2d(t/N)}}, \quad -\frac{\pi}{2} < \lambda < \frac{\pi}{2}. \tag{5.12}$$

As the process $y_{t,N}$ is non-stationary, increasing the number of observations by measuring new realizations of the process tells us nothing about the process' behavior at the beginning of the period. As a result, we fix the time period and we liken it to measuring the process at higher and higher levels of resolution on a fixed time interval along with the increase of N.

The spectral density for the process (5.11) is an even, 2π-periodic function that is uniformly Lipchitz continuous in $t/N \in [0,1]$. The time-varying spectral density function is given by:

$$f_N(\lambda) \sim (2|\lambda - \cos^{-1}\nu|)^{-2d(t/N)}, \quad \text{as} \quad \lambda \to \cos^{-1}\nu.$$

Then, if $d(t/N) > 0$, $f_N(\lambda)$ is smooth for frequencies around $\cos^{-1}\nu$, but is unbounded when $\lambda \to \cos^{-1}\nu$. In other words the behavior of $y_{t,N}$ is concentrated over the frequency associated with seasonality. This behavior can be extended to the case where we have several explosions inside the spectral density. This means that, on the interval $[0, N]$, we observe the locally stationary k-factor Gegenbauer process:

$$\prod_{i=1}^{k} (I - 2\nu_i B + B^2)^{d_i(t/N)} y_{t,N} = \varepsilon_t, \tag{5.13}$$

where ε_t is Gaussian white noise with the time-varying spectral density:

$$f_N(\lambda) = \frac{\sigma^2_{\varepsilon_t}}{2\pi} \prod_{i=1}^{k} (2|\cos\lambda - \nu_i|)^{-2d_i(t/N)}. \tag{5.14}$$

For the locally stationary 1-factor Gegenbauer process defined in (5.11), its time-varying spectral density is expressed in Equation (5.12). Thus by applying the logarithmic transform to both sides of Equation (5.12), we get

$$\log f_N(\lambda) = C - 2d(\frac{t}{N}) \log 2|\cos\lambda - \nu|, \quad -\frac{\pi}{2} < \lambda < \frac{\pi}{2}. \tag{5.15}$$

This suggests a simple regression of $\log f_N(\lambda)$ on $\log 2|\cos\lambda - \nu|$ to estimate the fractional difference parameter function $d(\frac{t}{N})$. For convenience, we approximate (5.15) by the following relationship:

$$\log f_N(\lambda) \approx C - 2d(\frac{t}{N}) \log 2|\lambda - \cos^{-1}(\nu)|, \quad -\frac{1}{2} < \lambda < \frac{1}{2}. \tag{5.16}$$

Thus, we partition the time interval $[0, 1)$ into 2^l $(0 < l < J-1)$ non-overlapping subintervals as follows:

$$I_h = [h2^{-2l}, (h+1)2^{-2l}), \quad h = 0, \cdots, 2^l - 1.$$

On each subinterval, we suppose that the time-varying parameter $d(\frac{t}{N})$ is locally constant, i.e., the process $(y_{t,N})_t$ is locally stationary on each subinterval $I_h = [h2^{-2l}, (h+1)2^{-2l}), h = 0, \cdots 2^l - 1$. Since the time-varying wavelet variance provides an estimate of the spectral density function, the logarithmic transformation of the variance of the wavelet coefficients provides the following log-linear relationship, for $i = 1, \cdots, k$:

$$\log \sigma_i^2(\lambda_{j,n}, t) = \alpha_i(t) + \beta_i(t) \log 2|\cos \mu_{j,n} - \nu_i|, \tag{5.17}$$

where $\sigma_i^2(\lambda_{j,n}, t)$ is the variance of the DWPT coefficients $\mathbf{W}_{j,n}$ associated, at time t, with the frequency interval $\lambda_{j,n} = (\frac{n}{2^{j+1}}, \frac{n+1}{2^{j+1}}]$ (where $n = 0, \cdots, 2^j - 1$; $j = 0, \cdots, J-1$); $\mu_{j,n}$ is the midpoint of the interval $\lambda_{j,n}$; $\beta_i(t)$ is the slope of the log-linear relationship at time t. Denote that $d_i(t) = -\beta_i(t)/2$. Now we apply locally the following approximation of the Equation (5.17) in order to estimate $d_i(t)$:

$$\log \sigma_i^2(\lambda_{j,n}, t) = \alpha_i(t) + \beta_i(t) \log 2|\mu_{j,n} - \cos^{-1}(\nu_i)| + u_i(t), \tag{5.18}$$

where $u_i(t)$ is a sequence of correlated random variables (we follow Arteche and Robinson (2000)'s methodology).

5.3.3 Procedure for Estimating $d_i(t)$ $(i = 1, \cdots, k)$

Using the previous filtering, we present a general procedure for estimating the time-varying parameter functions $d_i(t)$ $(i = 1, \cdots, k)$, of the model (5.13). We assume that the sample size is dyadic $(N = 2^J)$. We detail the different steps in order to estimate $d_i(t)$.

1. We first detect the Gegenbauer frequency λ_1 which corresponds to the highest explosion in the periodogram: $\nu_1 = \cos(\lambda_1)$. This frequency is fixed all along the procedure.

2. We compute the DWPT coefficient vectors $\mathbf{W}_{j,n}$ of length N_j through the formula (3.14), where $j = 0, \cdots, J-1$; $n = 0, \cdots, 2^j - 1$.

3. We associate to the vector $\mathbf{W}_{j,n}$ an adaptive orthonormal basis $\mathcal{B}$, such that the squared gain function of the wavelet filter associated with $\mathbf{W}_{j,n}$ is sufficiently small at the Gegenbauer frequency. Practically, we define $\mathcal{U}_{j,n}(f) = |U_{j,n}(f)|^2$ to be the squared gain function for the wavelet packet filter $u_{j,n,l}$, where $U_{j,n}(f)$ is the discrete Fourier transform (DFT) of

$$u_{j,n,l} = \sum_{k=0}^{L_j-1} u_{n,k} u_{j-1,[\frac{n}{2}],l-2^{j-1}k},\ l = 0, \cdots, L_j - 1,$$

with $u_{1,0,l} = g_l$, $u_{1,1,l} = h_l$ and $L_j = (2^j - 1)(L - 1) + 1$, g_l and h_l being the scaling filter and the wavelet filter defined as before.

4. The basis selection procedure involves selecting the combination of wavelet basis functions such that $\mathcal{U}_{j,n}(f_1) < \epsilon$ for some $\epsilon > 0$ at the minimum level j. However, the method of basis selection is not unique and the basis is not unique either. We apply the white noise tests like the portmanteau test to determine the best adaptive orthonormal basis that decorrelates the observed time series.

5. We partition the sampling interval $[0,\ 1)$ into 2^l non-overlapping subintervals of equal length, where l is an integer chosen such that $0 < l < (J - 1)$. "l" depends on the length of the data and the required precision. The 2^l subintervals are as follows

$$I_h = [h2^{-l}, (h + 1)2^{-l}),\ \ where\ \ h = 0, \cdots, 2^l - 1.$$

6. We locate the DWPT coefficients $W_{j,n,K}$ on each subinterval I_h. In order to construct the local estimates for the time-varying long memory parameter $d_1(t)$, we proceed according to the Heisenberg uncertainty principle: every DWPT coefficient vector is mapped to a rectangle (Heisenberg box) defined in the time-frequency plane with the boxes completely covering the plane.

7. Since the DWPT coefficient vector $\mathbf{W}_{j,n} = (W_{j,n,K})$ corresponds to the frequency interval $\lambda_{j,n} = (\frac{n}{2^{j+1}}, \frac{n+1}{2^{j+1}}]$, we obtain the corresponding time interval on the time-frequency plane with the width of $\mathbf{W}_{j,n}$ is $2^j/N$. Whereas, the length of the vector $\mathbf{W}_{j,n}$ is N_j. Therefore, we partition the elements of the vectors $\mathbf{W}_{j,n}$ with equal length $N_j/2^l = 2^{J-j-l}$ and attach them sequently to the subintervals I_h.

8. On each subinterval I_h $(h = 0, \cdots, 2^l - 1)$, we consider the bivariate collection of data

$$\{(\log 2|\mu_{j,n} - \cos^{-1}(\nu_1)|, \log \sigma_1^2(\lambda_{j,n})) \mid 0 \le n \le 2^j - 1; 0 \le j \le J - 1\},$$

then we get the approximate log-relationship (5.18). On each subinterval I_h, we carry out the ordinary least squares (OLS) regression to get the local estimates for the slope $\beta_1(t)$. Thus we obtain 2^l local estimates for the parameter $\beta_1(t)$. Since $\tilde{d}_1(t) = -\frac{\tilde{\beta}_1(t)}{2}$, we get 2^l local estimates for the parameter $d_1(t)$.

9. We omit the first and the last estimates to avoid the boundary effects. We associate the time index for $\tilde{d}_1(t)$ to the interval I_h by the midpoint of I_h, i.e. $2^{-l-1}(2h+1)$, and we smooth the estimated 2^l points by two local polynomial methods: spline method and loess (locally weighted scatter plot smoothing) method. Thus, we obtain two smoothed curves $\hat{d}_1(t)$ from the local estimates which approximate the true parameter curve.

10. The above steps (1-9) permit to get the estimate of $\hat{d}_1(t)$ and $\hat{\nu}_1$ for the corresponding Gegenbauer frequency $\hat{\lambda}_1$.

11. Now, we proceed in the same way to estimate the other Gegenbauer frequencies and their corresponding time-varying long memory parameters. First we calculate $y^1_{t,N} := (I - 2\hat{\nu}_1 B + B^2)^{\hat{d}_1(t)} y_{t,N}$, where $\hat{d}_1(t)$ and $\hat{\nu}_1$ are obtained in the previous steps. We need to interpolate some points such that the vector $\hat{d}_1(t)$ is of length N, due to the fact that the number of points on the smoothed curves is less than N if we adopt the loess smoothing method, for instance.

12. We repeat the above steps 1 to 9 on the vector $y^1_{t,N}$, then we get the estimate $\hat{d}_2(t))$ and $\hat{\nu}_2$ associated to the frequency $\hat{\lambda}_2$.

13. We proceed in the same way for other Gegenbauer frequencies until the $(k+1)$-th step providing the white noise $(\varepsilon_t)_t$.

At the end, there is no more peak in the periodogram, and we have k pairs of estimations for the Gegenbauer frequencies and parameter functions.

5.3.4 Consistency for Estimates $\hat{d}_i(t)$ ($i = 1, \cdots, k$)

In this subsection, we study the properties of the estimates $\hat{d}_i(t)$ ($i = 1, \cdots, k$). To get $\hat{d}_i(t)$, we have previously established a linear regression between the variance of the DWPT wavelet coefficients $W_{j,n,t}$ and the long memory parameters $d_i(t)$. Some similar approaches have been developed, in a stationary setting, by Geweke and Porter-Hudak (1983), Robinson (1995), Hurvich and Beltrao (1993) and Arteche (1998). And they obtain the consistency of the constant long memory parameter d_i. We will follow here the same method.

We have introduced the spectral density for the process $(y_{t,N})_t$ in (5.14). We will assume that the assumptions $\mathbf{A_1}$ - $\mathbf{A_2}$ and $\mathbf{A_4}$ - $\mathbf{A_5}$ introduced in Arteche (1998) are verified for $f_N(\lambda)$ defined in (5.14). The assumptions $\mathbf{A_1}$ and $\mathbf{A_2}$ specify the local behavior of the spectrum. The assumption $\mathbf{A_4}$ corresponds to the "trimming" condition introduced first in Robinson (1995). Now, under these assumptions and in the case of a 1-factor stationary Gegenbauer model, the asymptotic normality of the long memory parameter is obtained.

Lemma 5.3.2. *Consider the Gegenbauer model* $(I - 2\nu_1 B + B^2)^{d_1} y_t = \varepsilon_t$ *with the previous assumptions* $\mathbf{A_1}$ - $\mathbf{A_2}$ *and* $\mathbf{A_4}$ - $\mathbf{A_5}$. *Let* $\hat{d}_1$ *be the least squares estimate of* d_1

obtained from the following regression:

$$\log I(\omega+\lambda_j) = c + d_1(-2\log\lambda_j) + u_j, \;\; j = l+1, \cdots, m, \tag{5.19}$$

where $c = \log C - \eta$, η *is the Euler's constant* $\eta = 0.5772\cdots$,
$u_j = \log(\frac{I(\omega+\lambda_j)}{C\lambda_j^{-2d_1}}) + \eta$, $\lambda_j = \frac{2\pi j}{n}$ *are Fourier frequencies and* $I(\lambda)$ *is the periodogram. Then*

$$2\sqrt{m}(\hat{d}_1 - d_1) \rightarrow^d N(0, \frac{\pi}{6}).$$

This lemma has been proved by Arteche (1998).

Now we consider the locally stationary $1-$factor Gegenbauer process $(I-2\nu B+B^2)^{d_1(t)}y_t = \varepsilon_t$. The parameter $d(t)$ has been locally estimated on a sequence of subintervals I_h ($h = 0, \cdots, 2^l - 1$) treating locally as the stationary process, then we get

$$\forall h, \; \sqrt{m}(\hat{\beta}_1(h) - \beta_1(h)) \rightarrow^d N(0, \frac{\pi}{6}), \; h = 0, 1, \cdots, 2^l - 1.$$

Since $\tilde{d}_1(t) = -\frac{\tilde{\beta}(t)}{2}$, we have

$$\forall h, \; \sqrt{m}(\tilde{d}_1(h) - d_1(h)) \rightarrow^d N(0, \frac{\pi}{24}), \; h = 0, 1, \cdots, 2^l - 1.$$

In order to get a smoothed curve for $\hat{d}_1(t)$, we have smoothed the 2^l independent estimates $(\tilde{d}_1(0), \cdots, \tilde{d}_1(2^l-1))$ using two local polynomial methods: spline method and loess method, i.e., there exists a set of basis function $\hat{\omega}_h(t)$ satisfying that there exists a constant C such that $(\sum_{h=0}^{2^l-1} \omega_h(t)^2 = C < \infty)$ and

$$\hat{d}_1(t) = \sum_{h=0}^{2^l-1} \hat{\omega}_h(t)\tilde{d}_1(h).$$

Thus $E[\hat{d}_1(t)] = \sum_{h=0}^{2^l-1} \hat{\omega}_h(t)E[\tilde{d}_1(h)] = 0$, and $Var[\hat{d}_1(t)] = Var[\sum_{h=0}^{2^l-1} \hat{\omega}_h(t)\tilde{d}_1(h)] = (\sum_{h=0}^{2^l-1} \hat{\omega}_h(t))^2 Var[\tilde{d}_1(h)] = C\frac{\pi}{24} \equiv C_1$. Assuming that N tends to infinity, thus l tends to infinity, we can apply the central limit theorem and get the following result:

$$\sqrt{m}(\hat{d}_1(t) - d_1(t)) \rightarrow^d N(0, C_1).$$

This result is still true for all the time-varying parameters $d_i(t)(i = 1, \cdots, k)$, of a locally stationary k-factor Gegenbauer process.

5.4 Simulation Experiments

In this section, we carry out some Monte Carlo simulations to establish the robustness of the estimation of the parameter function $d_i(t)$ using wavelet approach, for finite samples. We focus on the model (5.10) with $(\varepsilon_t)_t$, a Gaussian noise:

$$(I - 2\nu B + B^2)^{d(t)} y(t) = \varepsilon(t). \tag{5.20}$$

We consider the constant, linear, quadratic, cubic, exponential and logarithmic functions $d(t)$ as follows:

1. $d(t)$ is constant: $d_0(t) = 0.3$;
2. $d(t)$ is linear: $d_1(t) = 0.2t + 0.1$;
3. $d(t)$ is quadratic: $d_2(t) = 0.3(t - 0.5)^2 + 0.1$;
4. $d(t)$ is cubic: $d_3(t) = 0.4(t - 0.5)^3 + 0.3$;
5. $d(t)$ is exponential: $d_4(t) = 0.01 \exp(2t) + 0.3$;
6. $d(t)$ is logarithmic: $d_5(t) = 0.1(\log(10t + 0.5) + 1.0)$.

For convenience, we assume that the observed data points $[y_1, \cdots, y_N]^T$ are equally spaced on the time interval and are scaled on the time interval $[0, 1)$, using the transformation $t_i = \dfrac{i-1}{N}$ (where $i = 1, \cdots, N = 2^J$). In our examples, we use $N = 4096 = 2^{12}$ ($J = 12$) and the Gegenbauer frequency $cos^{-1}\nu = \lambda_G = \dfrac{\pi}{3}$.

For the estimation procedure, we use the MB(16) wavelet filter ($L = 16$), and we choose the adaptive orthonormal basis using portmanteau test with $p = 0.01$. We partition the sampling interval $[0, 1)$ into $2^6 = 64$ subintervals ($l = 6$) and, we get 64 local estimates for $d(t)$. Finally, we smooth the estimates using two local polynomial methods, spline method and loess method. We replicate the simulations 100 times for each locally stationary 1-factor GARMA process (5.10) with the two previous functions $d(t)$. We carry out the code on the computer Mac OS X 10.5.1 Léopard, written in language R with the help of the package "waveslim".

We denote $(y_{0,t})_t, \cdots, (y_{5,t})_t$, the process (5.10) for constant, linear, quadratic, cubic, exponential and logarithmic parameter functions expressed in the beginning of this part. For each process, we provide the trajectories, the autocorrelation function, spectrum and the estimated curves smoothed by spline method and loess method together with the true functions of $d(t)$, see in from Figure 5.1 to Figure 5.24 for more details.

In Table 5.1, numerical results for the six models $(y_{0,t})_t$, $(y_{1,t})_t$, $(y_{2,t})_t$, $(y_{3,t})_t$, $(y_{4,t})_t$ and $(y_{5,t})_t$ are provided with 100 times replication. We give the mean of the estimated Gegenbauer frequencies, mean of the bias and that of the RMSE for $\hat{d}(t)$ using 100 simulations. We deduce that:

	Gegenbauer frequency $\cos^{-1}\nu$	bias of $\hat{d}(t)$	RMSE of $\hat{d}(t)$
$y_{0,t}$	0.333403	spline: -0.159373 loess: -0.120514	spline: 0.174109 loess: 0.127925
$y_{1,t}$	0.33262	spline: -0.115436 loess: -0.089052	spline: 0.126721 loess: 0.095458
$y_{2,t}$	0.335110	spline: -0.006068 loess: -0.001749	spline: 0.039389 loess: 0.032570
$y_{3,t}$	0.333198	spline: -0.166090 loess: -0.123020	spline: 0.178187 loess: 0.130845
$y_{4,t}$	0.333235	spline: -0.169756 loess: -0.122901	spline: 0.185885 loess: 0.131963
$y_{5,t}$	0.333284	spline: -0.171724 loess: -0.133765	spline: 0.185447 loess: 0.138645

Table 5.1: Estimation of Gegenbauer frequencies, bias and RMSE of $(y_{0,t})_t$, $(y_{1,t})_t$, $(y_{2,t})_t$, $(y_{3,t})_t$, $(y_{4,t})_t$, $(y_{5,t})_t$.

1. Each estimated curve approximates the general shape of the time-varying parameter function. The rebuilding of the curve smoothed using the loess method appears better than that using the spline method.

2. The estimations of the Gegenbauer frequencies have quite small bias. The small values for the bias and the RMSE of the estimated parameter suggest that our algorithm is robust. Comparing the two smoothing methods, we find that in most cases, the loess method performs a little better than the spline method.

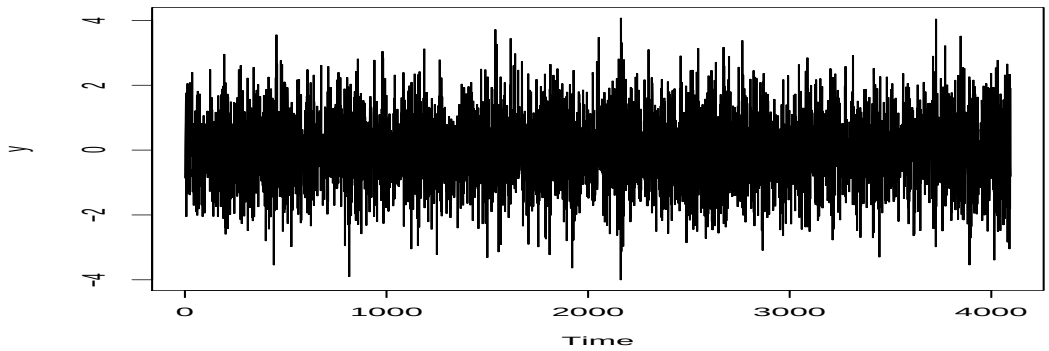


Figure 5.1: Sample path of $(y_{0,t})_t$

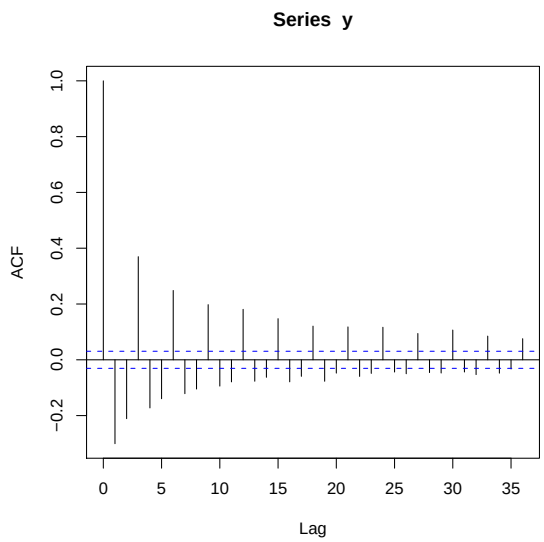


Figure 5.2: ACF of $(y_{0,t})_t$

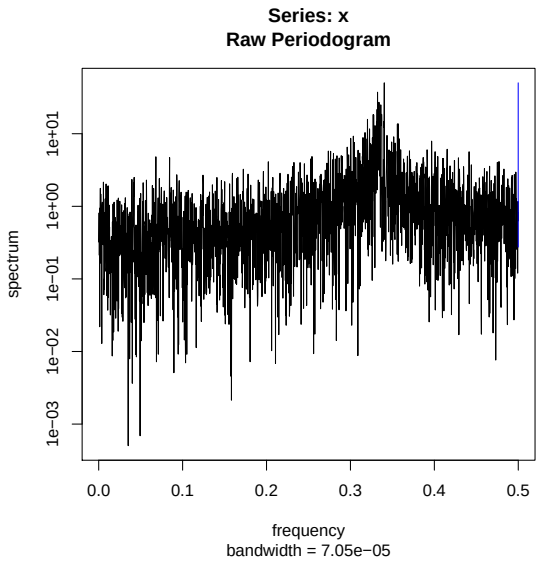


Figure 5.3: Spectrum of $(y_{0,t})_t$

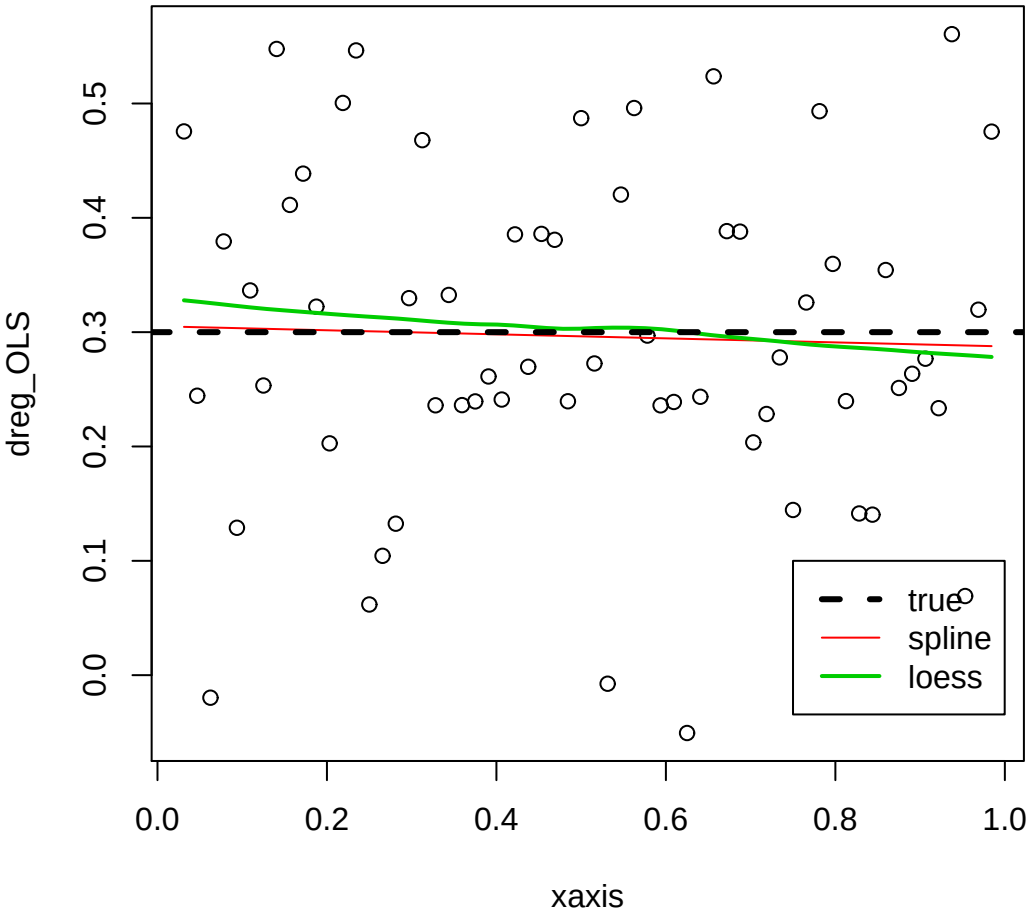


Figure 5.4: $\tilde{d}_0(t)$ (smoothed by spline and loess method) for $(y_{0,t})_t$

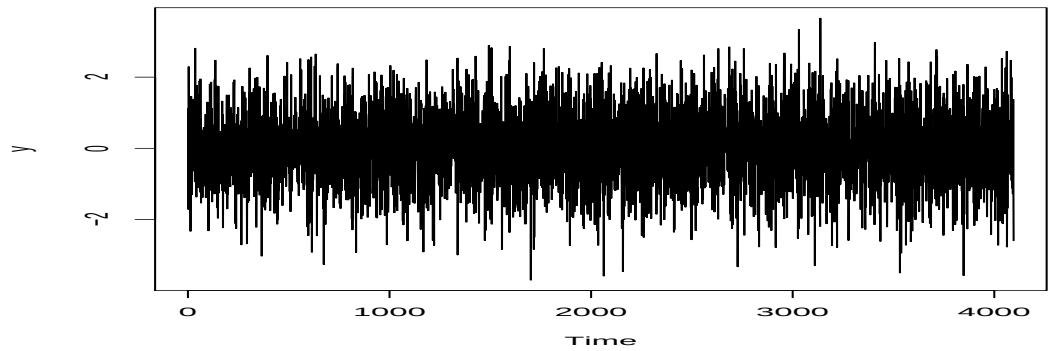


Figure 5.5: Sample path of $(y_{1,t})_t$

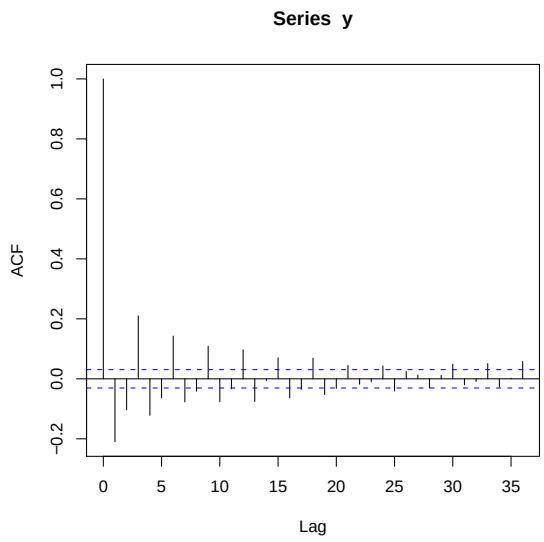


Figure 5.6: ACF of $(y_{1,t})_t$

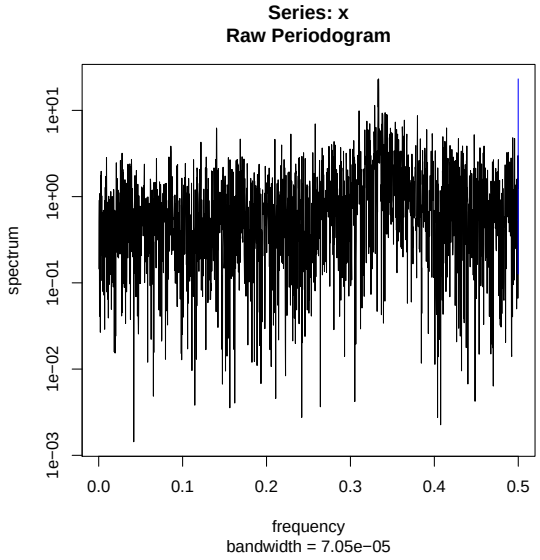


Figure 5.7: Spectrum of $(y_{1,t})_t$

5.5 Forecast for Non-stationary Processes

There exists much literature concerning the forecast for stationary processes, see for example, Geweke and Porter-Hudak (1983), Barkoulas and Baum (1997), Noakes et al. (1988), Ray (1993a), Smith and Yadav (1994), Crato and Ray (1996), Brodsky and Hurvich (1999), etc.

As for the forecast of non-stationary process, to our knowledge, there exists almost no literature for reference. In this thesis, since we proposed a new non-stationary process together with a consistent and robust wavelet-based estimation procedure, it is necessary to give out a method of the forecast for this new model.

It is known that the best prediction on a horizon h for a time series y_t is in the sense of the minimum mean squared error (MSE). The forecast provided by the predictor of least squares method is given by the following expression:

$$\hat{y}_t(h) = E(y_{t+h}|I_t), \;\; where \;\; I_t = \sigma(y_s, s \leq t).$$

We assess the predictive ability of the model by considering the root mean square error (RMSE) of prediction. The criteria are defined as follows:

$$RMSE = \sqrt{\frac{1}{h}\sum_{l=1}^{h}(y_{t+l} - \hat{y}_t(l))^2},$$

where h is the forecast horizon and $\hat{y}_t(l)$ is the predicted value of y_{t+l}, see Priestley (1981) for reference.

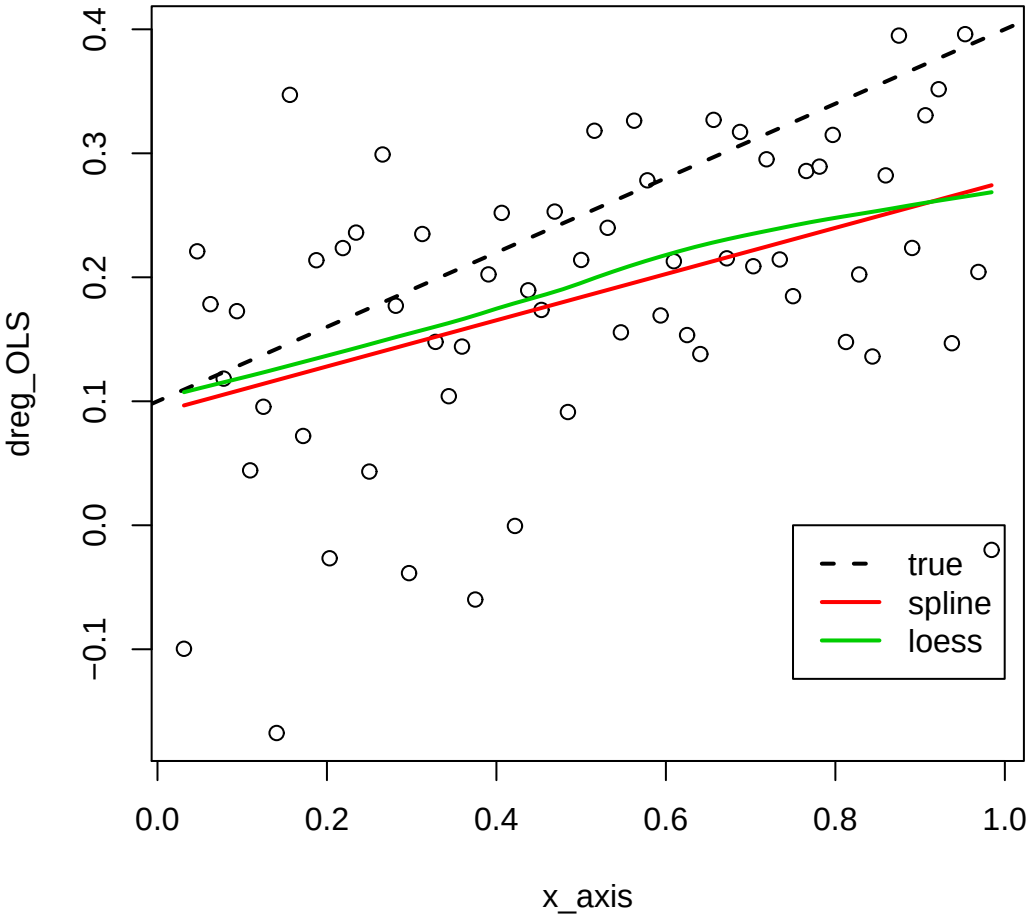


Figure 5.8: $\tilde{d}_1(t)$ (smoothed by spline and loess method) for $(y_{1,t})_t$

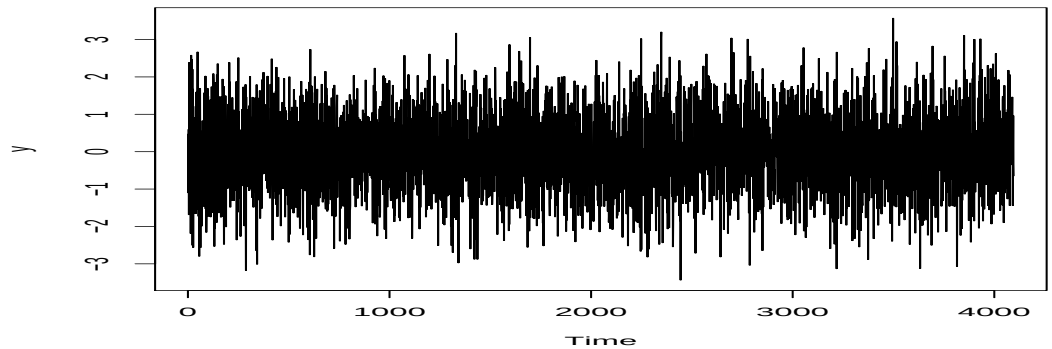


Figure 5.9: Sample path of $(y_{2,t})_t$

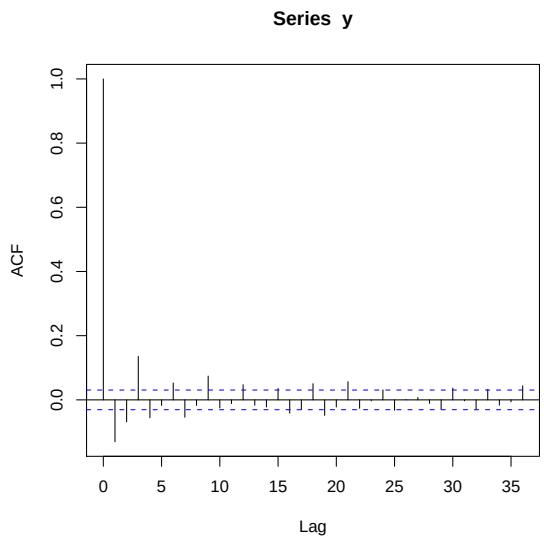


Figure 5.10: ACF of $(y_{2,t})_t$

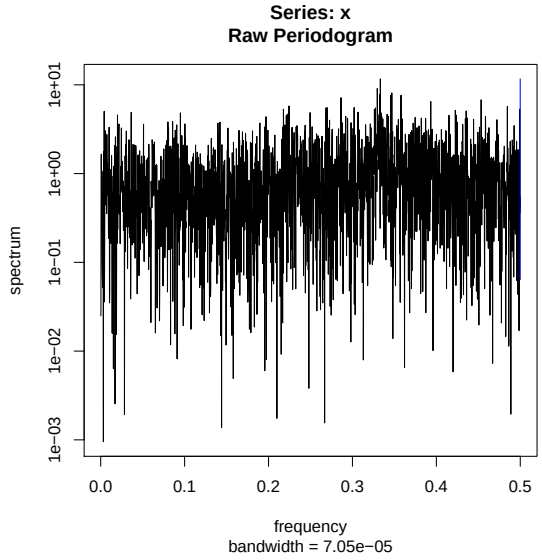


Figure 5.11: Spectrum of $(y_{2,t})_t$

Now, we consider the forecast for non-stationary processes, especially the locally stationary k-factor Gegenbauer process that we have proposed:

$$\prod_{i=1}^{k}(I - 2\cos\lambda_i B + B^2)^{d_i(t)} y_t = \varepsilon(t), \tag{5.21}$$

where $(\varepsilon_t)_t$ is white noise with variance σ_ε^2.

Our strategy for the prediction of the locally stationary k-factor Gegenbauer process is as follows:

1. Make the n-step-ahead forecast for the long memory parameter functions $\hat{d}_t(h) = E(d(t+h)|I_t)$, where $I_t = \sigma(d(s), s \leq t)$.
   We start from the estimations of $\hat{d}_i(t)$ obtained by the wavelet-based algorithm which are smoothed by spline method and loess method. It is the prediction for the estimated polynomial curves.

2. Calculate y_t using the Equations (5.5) and (5.6).

In detail, suppose that we have the series $y_1, \cdots, y_T$ which has been modeled by the locally stationary k-factor Gegenbauer process (5.21), and the estimated time-varying parameters are $\hat{d}_i(t)$, $t = 1, \cdots, T$. First we make the n-step-ahead forecasts $\hat{d}_i(T+h)$ $(h = 1, \cdots, n)$ for the time-varying long memory parameters. One smoothed curve is obtained by the cubic smoothing spline method. We predict a smoothing spline fit at new points. The predicted fit is linear beyond the original data. The other smoothed curve is obtained by the loess method using local fitting by fitting a polynomial surface determined by one or more numerical predictors. According to the results of Monte Carlo experiments, when the degree of smoothing is 0.75, the estimation behavior seems to

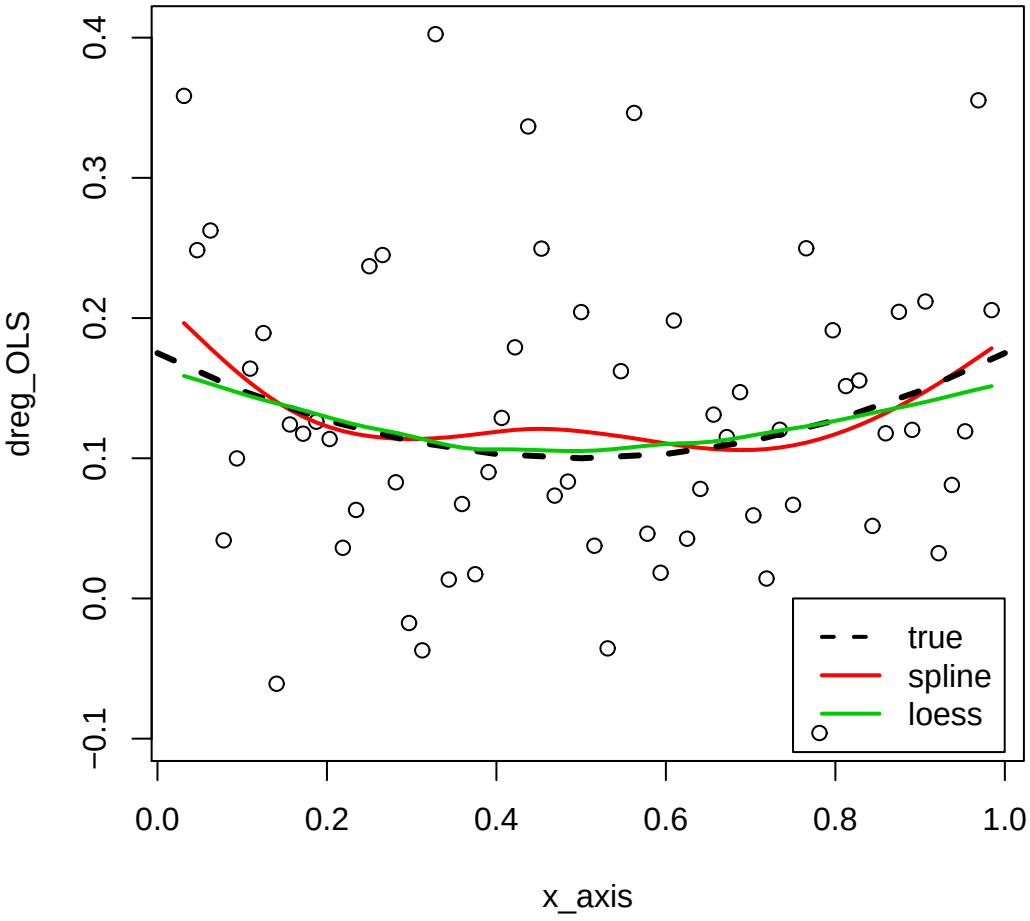


Figure 5.12: $\tilde{d}_2(t)$ (smoothed by spline and loess method) for $(y_{2,t})_t$

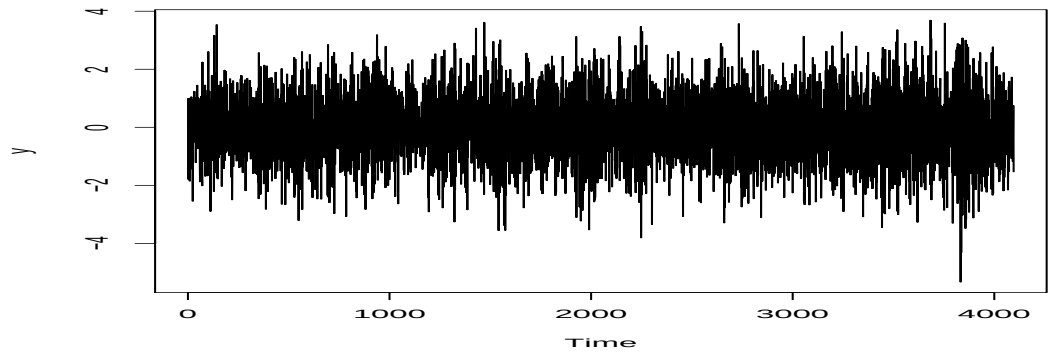


Figure 5.13: Sample path of $(y_{3,t})_t$

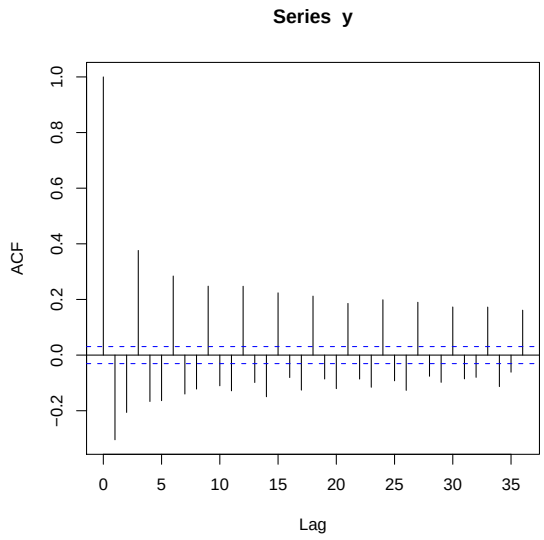


Figure 5.14: ACF of $(y_{3,t})_t$

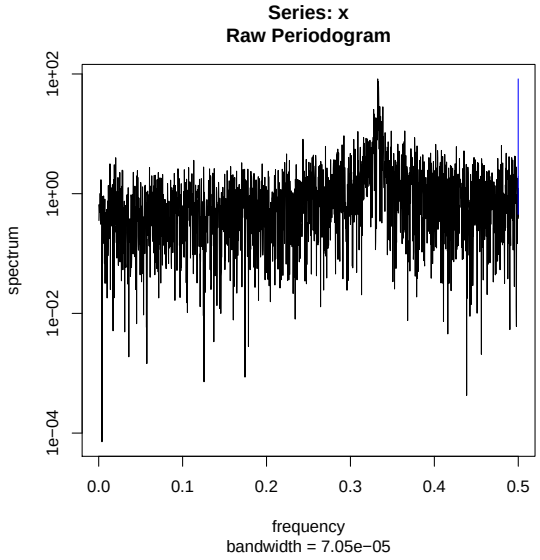


Figure 5.15: Spectrum of $(y_{3,t})_t$

be the best for the loess method with comparison to the other degrees of smoothing. And a few iterations of an M-estimation procedure with Tukey's biweight are used to achieve the forecast of $\hat{d}(t)$. Then we obtain the forecast $\hat{y}_{T+h}$ $(h = 1, \cdots, n)$ using the orthogonal Gegenbauer coefficients. To justify the behavior of the forecast, we measure the corresponding bias and RMSE.

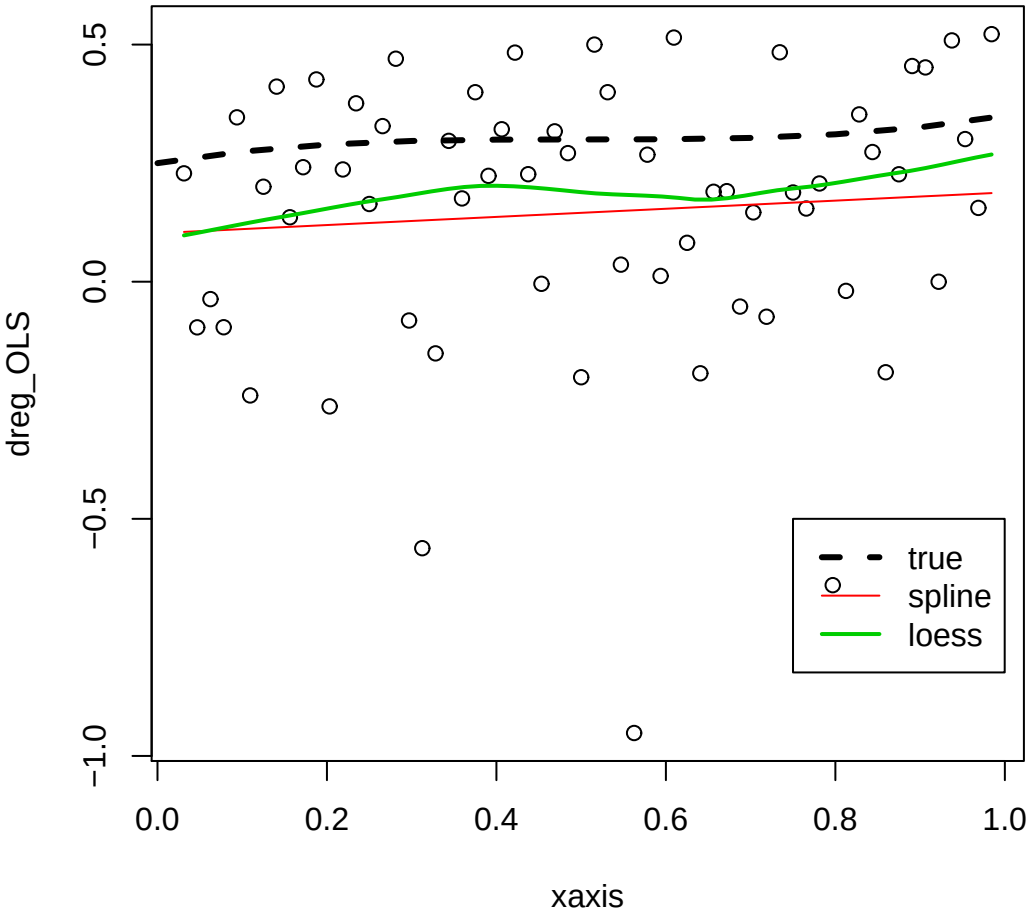


Figure 5.16: $\tilde{d}_3(t)$ (smoothed by spline and loess method) for $(y_{3,t})_t$

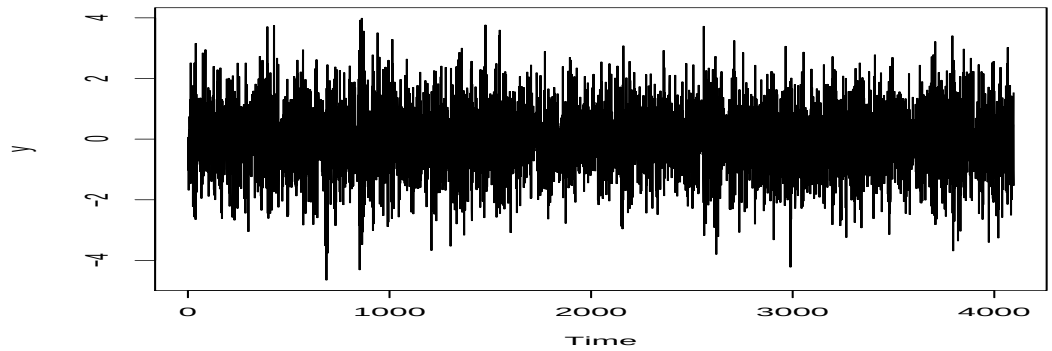


Figure 5.17: Sample path of $(y_{4,t})_t$

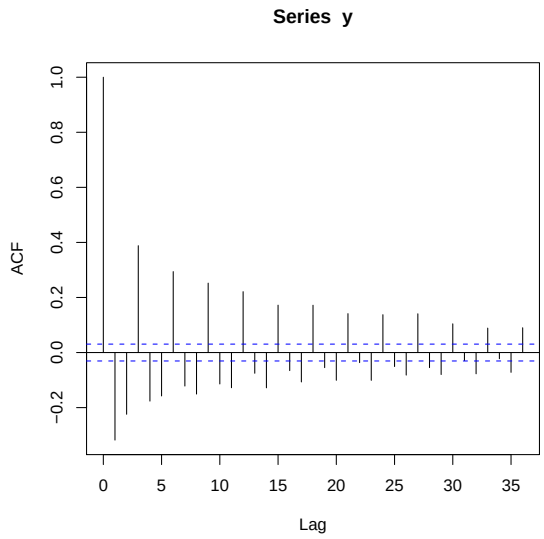


Figure 5.18: ACF of $(y_{4,t})_t$

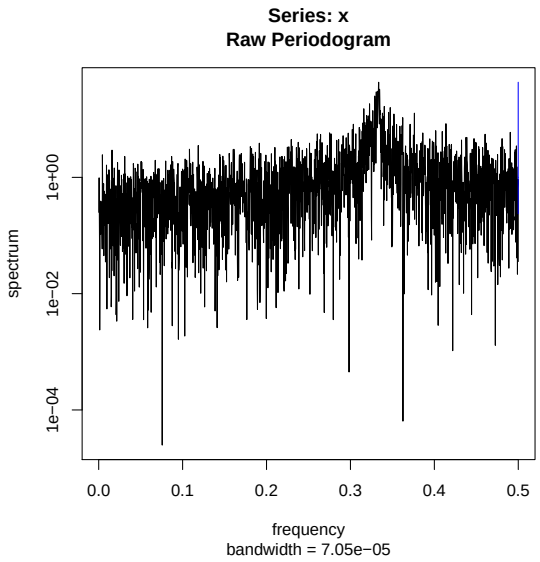


Figure 5.19: Spectrum of $(y_{4,t})_t$

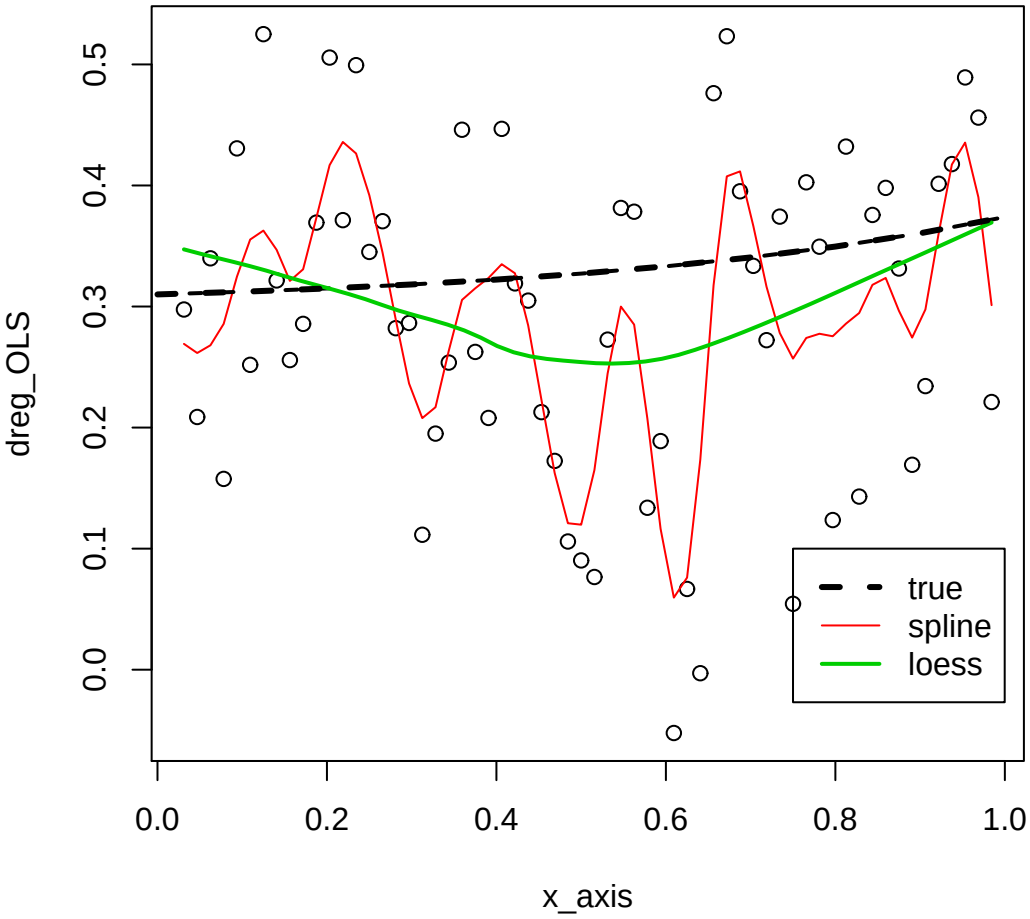


Figure 5.20: $\tilde{d}_4(t)$ (smoothed by spline and loess method) for $(y_{4,t})_t$

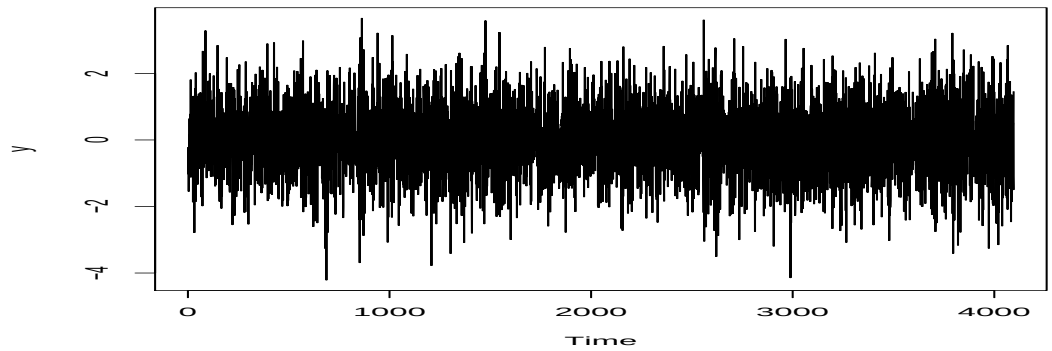


Figure 5.21: Sample path of $(y_{5,t})_t$

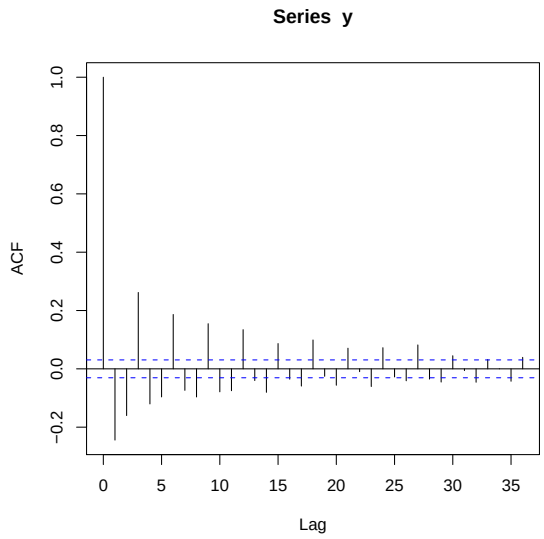


Figure 5.22: ACF of $(y_{5,t})_t$

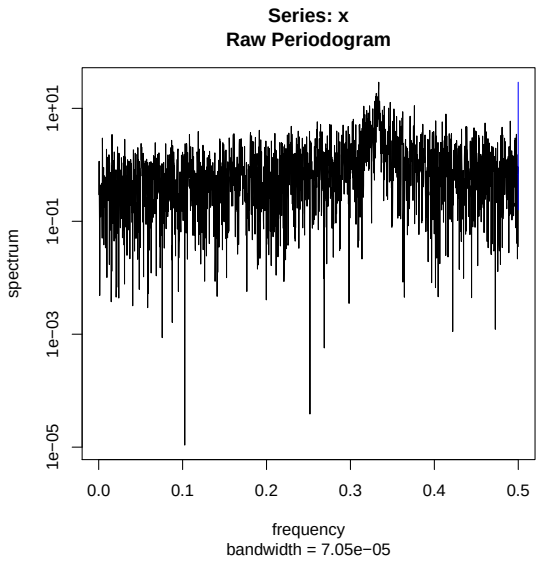


Figure 5.23: Spectrum of $(y_{5,t})_t$

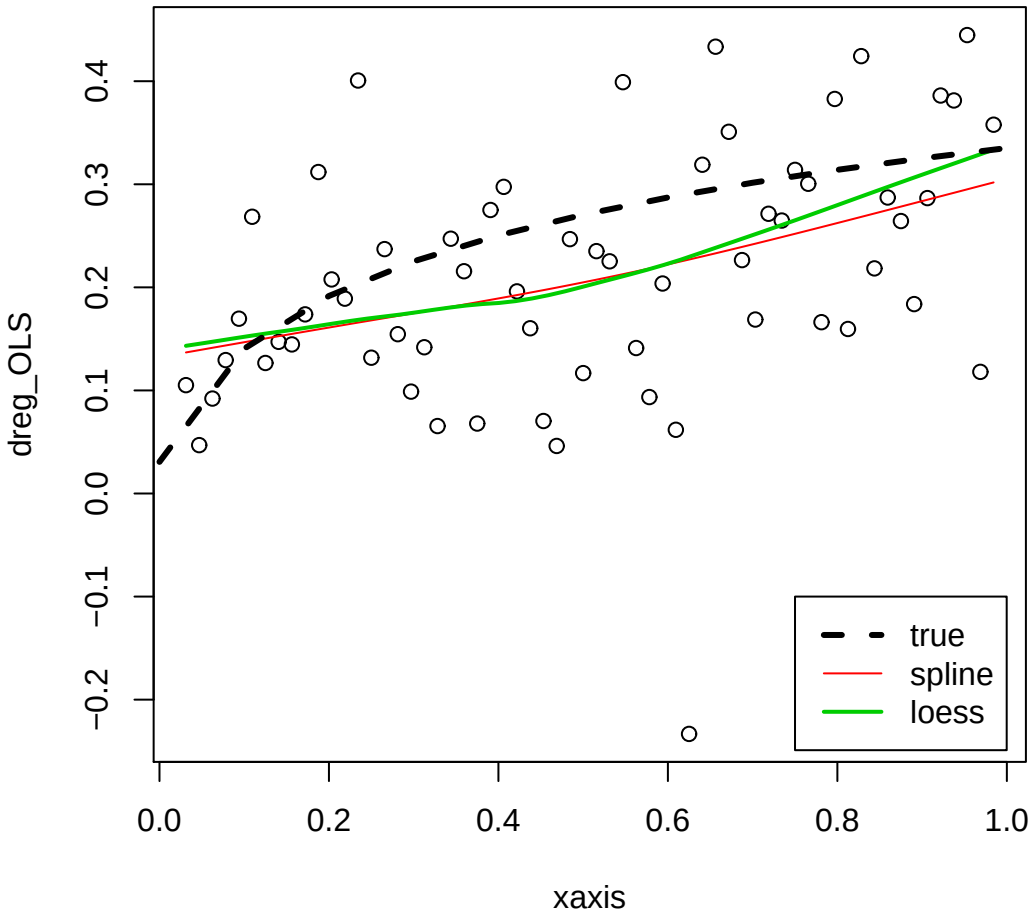


Figure 5.24: $\tilde{d}_5(t)$ (smoothed by spline and loess method) for $(y_{5,t})_t$

Chapter 6

Applications

This chapter is concerned with the application results of the financial and energy data, applying the estimation theory of stationary and non-stationary long memory processes. In particular, for these two series, we apply the new non-stationary model and wavelet-based estimation method that we proposed in Chapter 5. Then we make the predictions.

6.1 Nikkei Stock Average 225 Index Data

6.1.1 Data Set

Many authors have made the applications on the stock markets, for example, Dufrénot et al. (2005a) and Ferrara and Guégan (2001a). In this section we consider the Nikkei Stock Average 225 (NSA 225) spot index and futures price which correspond to 4096 daily observations of the spot index and the futures price of the NSA 225, covering the period from January 2nd 1989 to September 13rd 2004. Daily closing values of the spot index and the settlement prices of the futures contracts are used. The regular futures contracts mature in March, June, September and December. For further details on the futures price series, we refer to Lien and Tse (1999). Data sets are available from Thomson data stream.

6.1.2 Modeling

Figure 6.1 represents the daily spot index and daily futures prices from January 2nd 1989 to September 13rd 2004. $(S_t)_t$ denotes the logarithm of the spot price and $(F_t)_t$ the logarithm of the futures price. Lien and Tse (1999) assumed that $(S_t)_t$ and $(F_t)_t$ were both integrated of order one and they modeled the relationship between $(S_t)_t$ and $(F_t)_t$ using an error correction model (ECM), proposed by Engle and Granger (1987). Current prices are affected by the past prices and error correction term and the authors used the following relationship:

$$\Delta S_t = \phi_0 + \sum_{i=1}^{p} \phi_i \Delta S_{t-i} + \sum_{j=1}^{q} \psi_j \Delta F_{t-j} + \gamma Z_{t-1} + \varepsilon_{S_t}, \tag{6.1}$$

where $(Z_t)_t$ is such that $Z_t = F_t - S_t$, for $t = 1, \cdots, T$. Figure 6.2 represents the error correction term $(Z_t)_t$, which is the difference between the log futures prices and the log

spot prices.

Our aim is to model the error correction term $(Z_t)_t$ using the non-stationary model and the method that we have developed in Chapter 5. Indeed, Lien and Tse (1999) and Ferrara and Guégan (2001b) have already considered this problem using stationary models on a shorter period (from January 1989 to August 1997 and from May 1992 to August 1996 respectively). In the error correction model (ECM) of Lien and Tse (1999), the spot prices and futures prices are integrated of order one but the bias (the difference between the futures price and the spot indexes) is fractionally integrated. Whereas, Ferrara and Guégan (2001b) modeled the bias in the ECM by stationary Gegenbauer process, which has been proved to be more efficient than the modeling proposed by Lien and Tse (1999).

However, what we consider is an even longer time period of data set, which is not necessarily globally stationary. Since we observe the existence of volatility in $(Z_t)_t$, it seems appropriate to model the series $(Z_t)_t$ by non-stationary model with time-varying parameters. In the following, we propose to model the series using the locally stationary k-factor GARMA process.

In the first step, we use the wavelet multiresolution analysis to remove the time-varying mean. For this purpose, we apply the Maximal Overlap Discrete Wavelet Transform (MODWT) (J=6) with a Daubechies least asymmetric (LA(8)) wavelet filter and perform the multiresolution analysis, which is shown in Figure 6.3. The wavelet details, $D_1, \cdots, D_6$, exhibit the zero mean, while the wavelet smooth S_6, associated with the low frequency $[0, 1/64]$, captures the trend of the series. To remove the time-dependent mean, we ignore the wavelet smooth and sum up the six wavelet details. We get the residuals: $Z_t - S_{6,t}$ where $S_{6,t}$ is the wavelet smooth. Thus, the residuals still keep the periodicity in the original data.

In the second step, we estimate the Gegenbauer frequency $\hat{\lambda}_G = 0.015$ which corresponds to the highest explosion in the periodogram. Thus, we apply the DWPT on the residuals $Z_t - Z_{6,t}$, choosing the orthonormal basis, locating the DWPT coefficients on the evenly partitioned 64 subintervals, calculating locally the variance, and carrying out the approximate OLS regression on each subinterval. Then, we smoothed the local estimated points of the long memory parameter by spline and loess method, without taking into account the first and last estimated points to avoid the boundary effects.

Thus, we fit the residuals by the following model:

$$(I - 2 \times 0.995B + B^2)^{\hat{d}(t)}(Z_t - S_{6,t}) = \varepsilon_t, \tag{6.2}$$

where $\hat{d}(t)$ is the estimated curves presented in Figure 6.4, $\nu = 0.995 = cos(2\pi\hat{\lambda}_G)$ and $S_{6,t}$ is the wavelet smooth obtained by the multiresolution analysis indicated in Figure 6.3. The thin real curve is the estimated parameter function obtained by spline smoothing, and the thick real curve is smoothed by loess method. In Figure 6.4, we can also observe the estimation results in dashed line and dotted line using another two semi-parametric methods, Robinson method (1995) and Whittle method (1951), treating the parameter

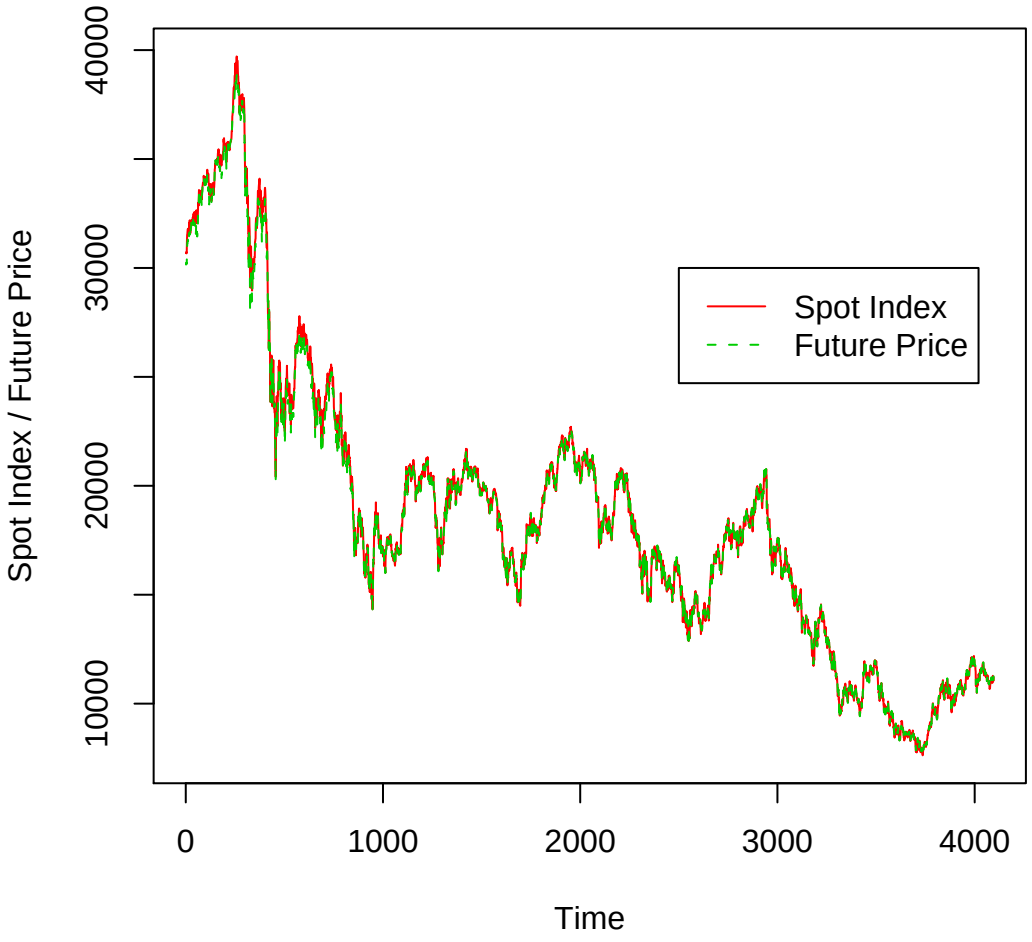


Figure 6.1: Trajectory of NSA 225 index (02/01/1989-13/09/2004)

function as a constant in the model 6.2. Comparing our result with that in Ferrara and Guégan (2001b) on the same time period, we get similar behavior, while our result capturing more precisely the local changes.

Therefore, the new model and the new method that we developed in Chapter 5 permit to extend the other authors' work by capturing the local characteristics, which indicates the advantages of our methodology. Indeed, on a shorter time period, it seems to be reasonable to assume the stationarity for the series. However on a longer time period, since general stationarity is not always satisfied, it seems to be more appropriate to consider the non-stationary processes and the time-varying parameters and to work locally.

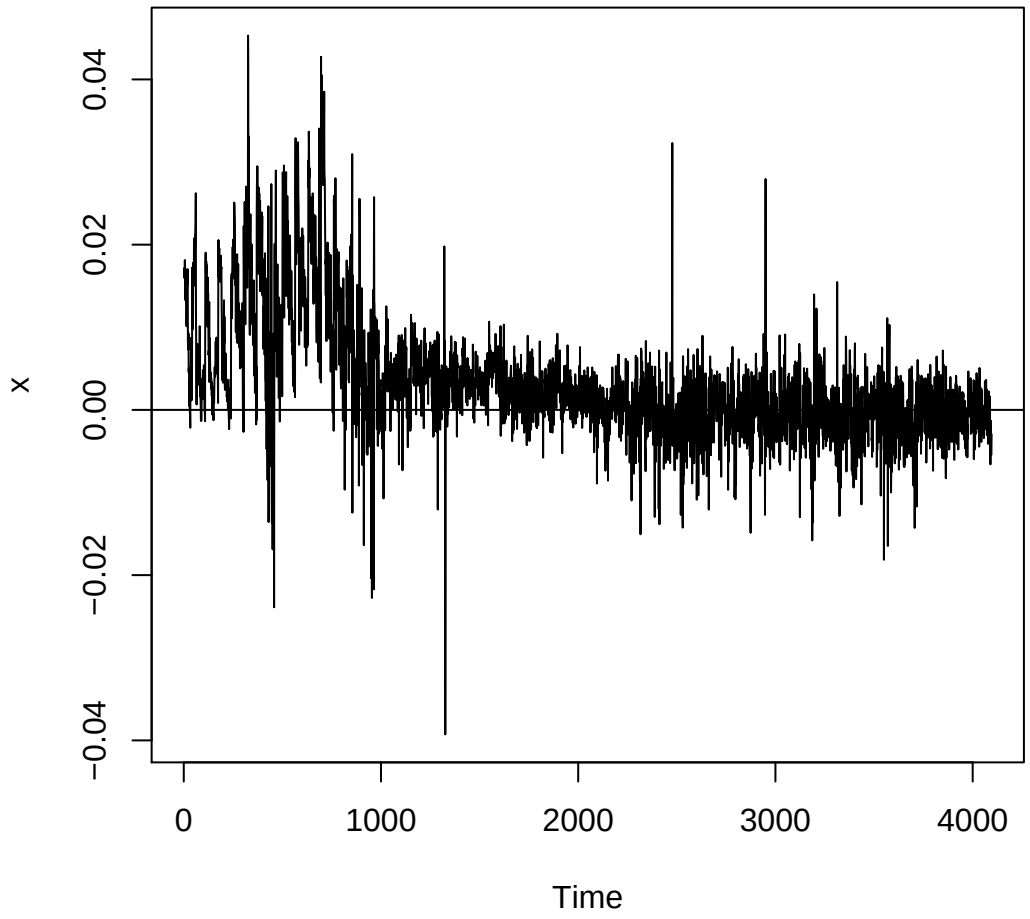


Figure 6.2: Error Correction Term $(Z_t)_t$

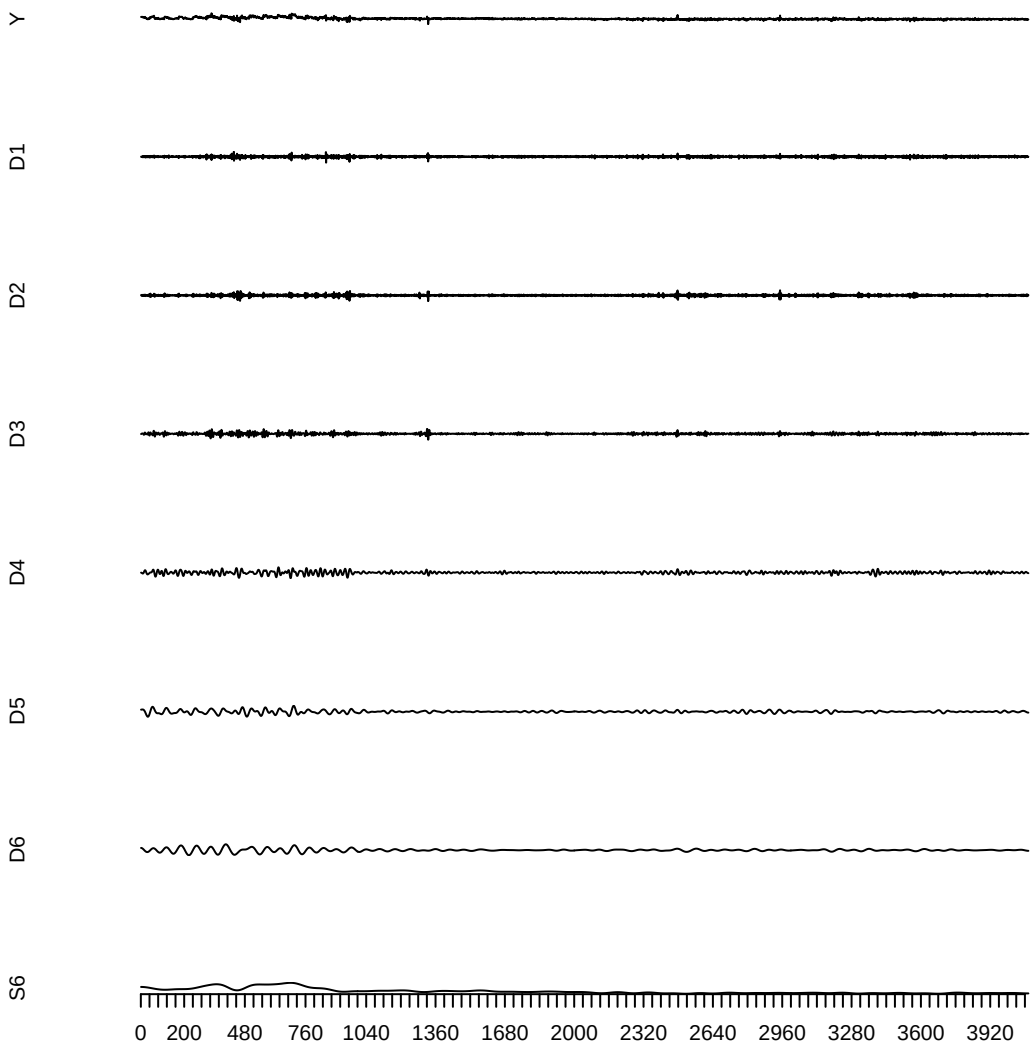


Figure 6.3: Multi-Resolution Analysis of $(Z_t)_t$ ($J = 6$)

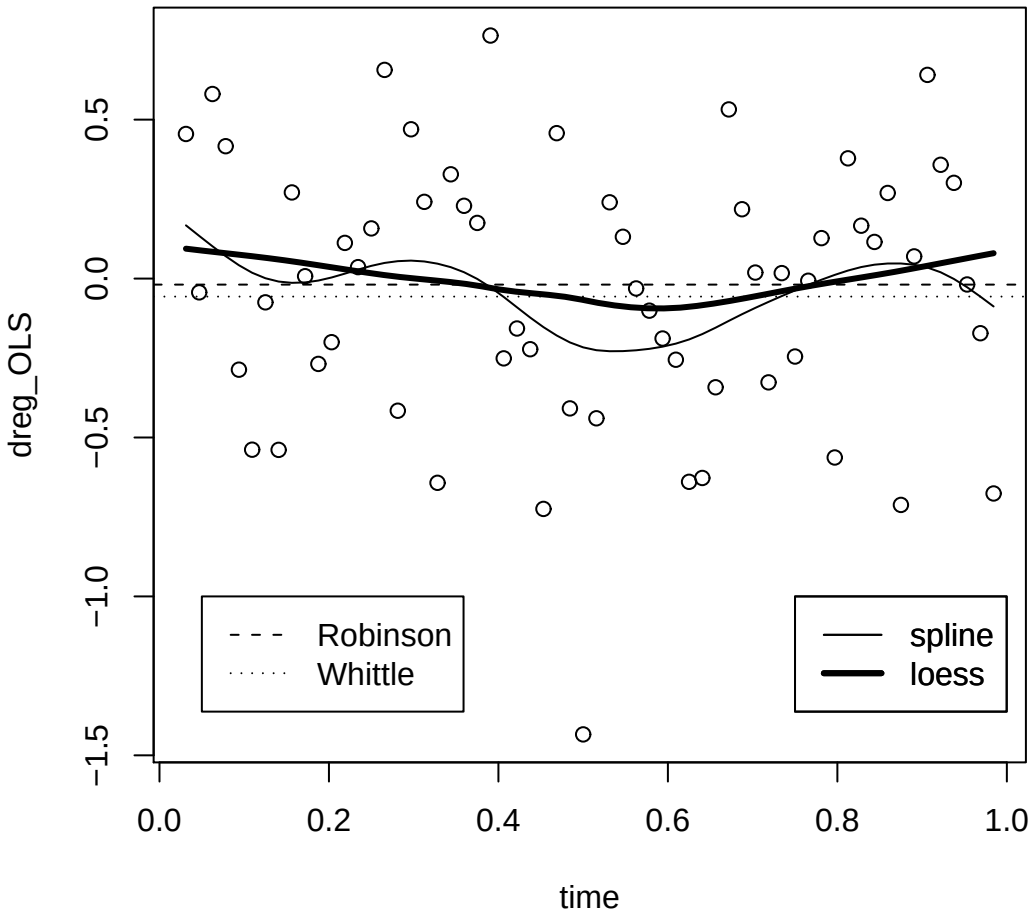


Figure 6.4: $\tilde{d}(t)$ smoothed by spline and loess method

	$h=1$	$h=2$	$h=3$	$h=4$	$h=5$	$h=6$	$h=7$
Bias	23.30081	23.06577	23.26697	22.83360	22.88856	23.30048	23.20274
RMSE	0.465407	0.3278092	0.2697851	0.2288531	0.2071427	0.1911432	0.1750827

Table 6.1: The relative results of the h-step-ahead predictions on the error correction term in the ECM of NSA 225 index data using the locally stationary 1-factor Gegenbauer model (parameter function smoothed by spline method).

6.1.3 Forecast

In the previous part, we have modeled the error correction term in the ECM of the NSA 225 index by the locally stationary 1-factor Gegenbauer process. Now we turn to the forecast of the model. According to the forecast theory described in Chapter 5, what is critical is to forecast the time-varying long memory parameter function $\hat{d}(t)$ in the model.

Using the strategy that we described in Chapter 5, first of all, we make the the h-step-ahead ($h = 1, \cdots, 7$) forecasts for $\hat{d}(t)$ smoothed by spline method and that by loess method. Once we obtain the h-step-ahead predictions for $\hat{d}(t)$, we can deduce the h-step-ahead predictions for Z_t. We present in the following the prediction results according to the smoothing method for the estimated long memory parameter functions.

- The first case: the parameter function is obtained by the spline smoothing.

  We first make the h-step-ahead forecasts for $\hat{d}(t)$ ($h = 1, \cdots, 7$). We can observe the predictions for $\hat{d}(t)$ in Figure 6.5 ($h = 1$), Figure 6.6 ($h = 2$), Figure 6.7 ($h = 3$), Figure 6.8 ($h = 4$), Figure 6.9 ($h = 5$), Figure 6.10 ($h = 6$) and Figure 6.11 ($h = 7$). Then we utilize these results to forecast the error correction term Z_t. Figure 6.12 ($h = 1$), Figure 6.13 ($h = 2$), Figure 6.14 ($h = 3$), Figure 6.15 ($h = 4$), Figure 6.16 ($h = 5$), Figure 6.17 ($h = 6$) and Figure 6.18 ($h = 7$) are the corresponding predictions for Z_t.

  From the graphs, there seems to exist some differences between the true trajectory (see in Figure 6.2) and the predicted trajectory. To our understanding, it is to some extent due to the different scales in the graphs. To have a more clear idea of our forecast, we need to refer to the criteria of the forecast. In Table 6.1, the bias and the RMSE of the h-step-ahead predictions are presented. We find that although Z_t is a long series with the length 4096, the bias and the RMSE are quite small, which is really satisfying.

- The second case: the parameter function is smoothed by the loess method.

  We first make the h-step-ahead forecast for $\hat{d}(t)$ ($h = 1, \cdots, 7$). We can observe the predictions for $\hat{d}(t)$ in Figure 6.19 ($h = 1$), Figure 6.20 ($h = 2$), Figure 6.21 ($h = 3$), Figure 6.22 ($h = 4$), Figure 6.23 ($h = 5$), Figure 6.24 ($h = 6$) and Figure 6.25 ($h = 7$). Then we utilize these results to forecast the error correction term Z_t, using the strategy described in the previous Chapter. In Figure 6.26, Figure 6.27,

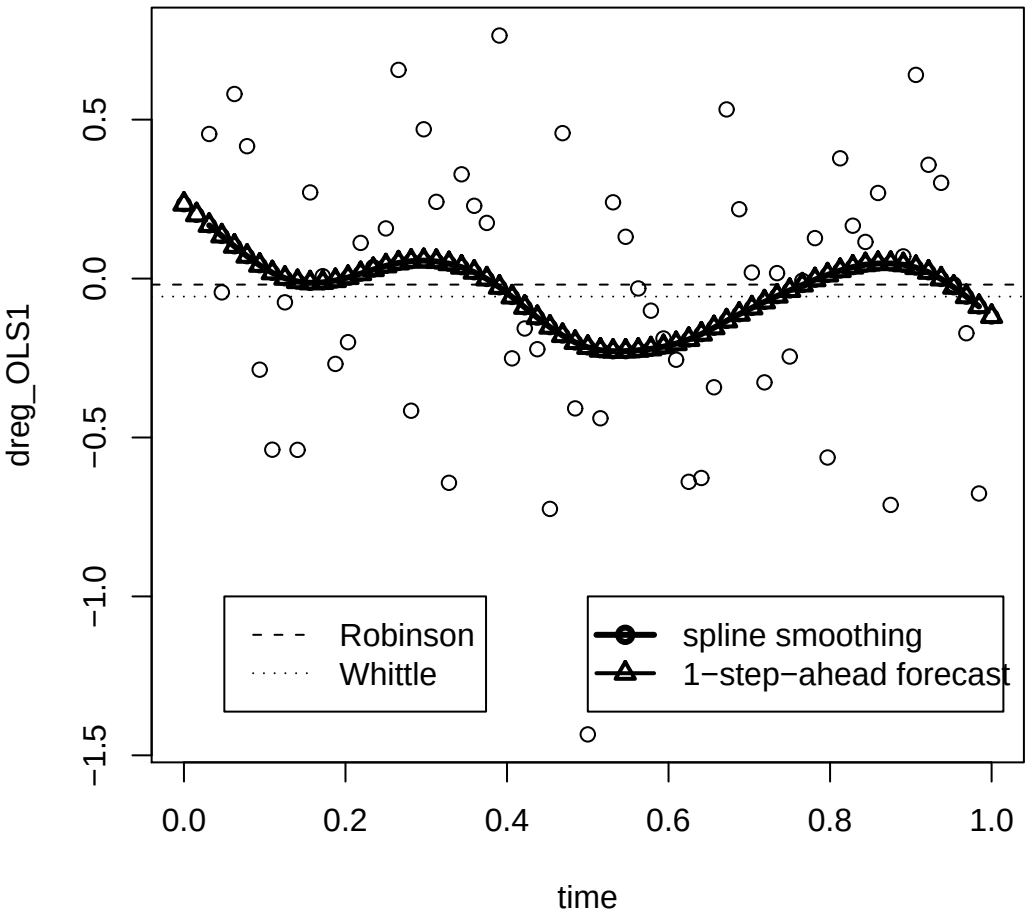


Figure 6.5: NSA: 1-step-ahead forecast for $\tilde{d}(t)$ smoothed by spline method

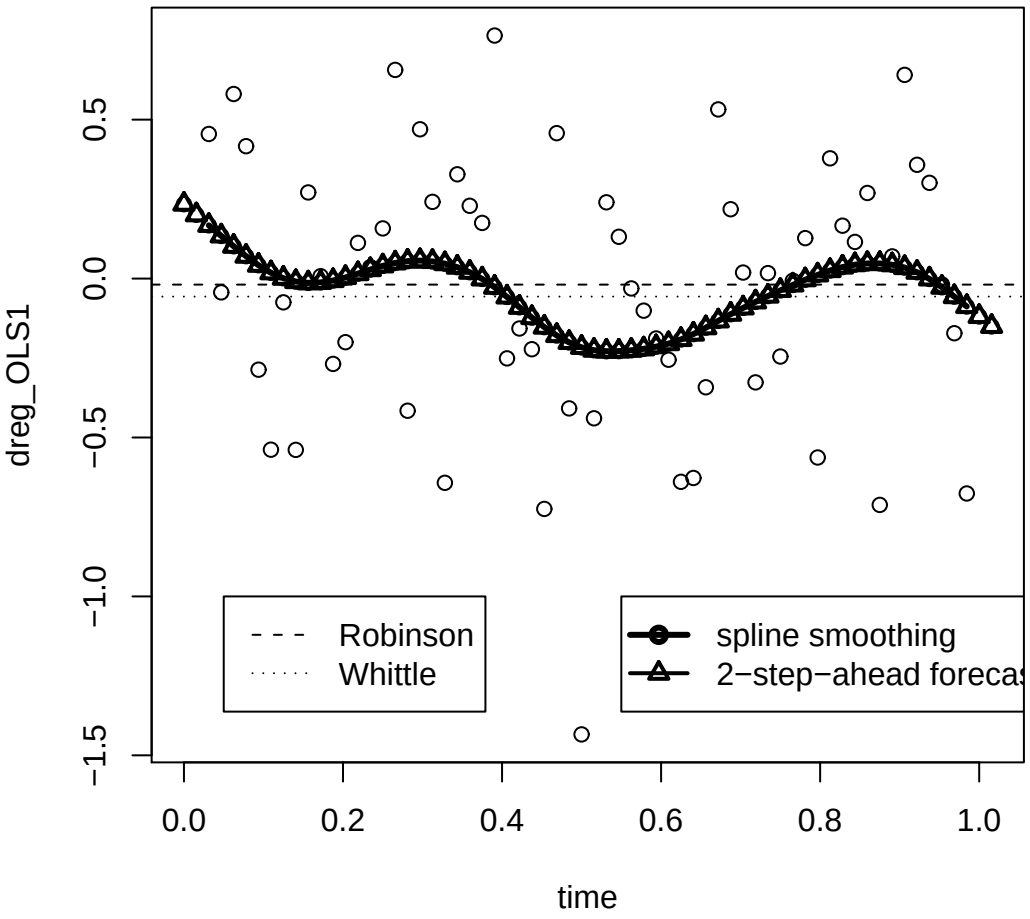


Figure 6.6: NSA: 2-step-ahead forecast for $\tilde{d}(t)$ smoothed by spline method

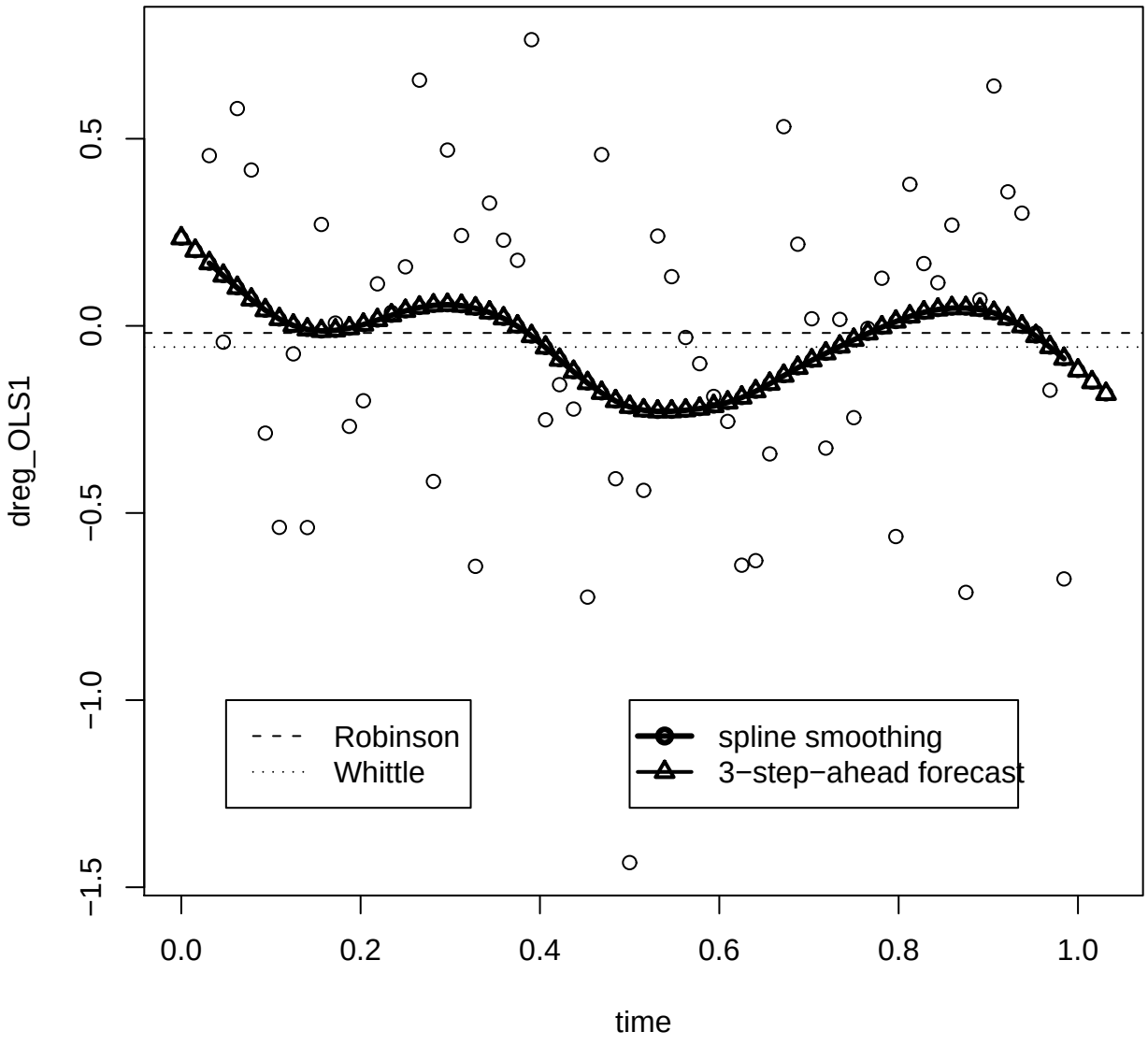


Figure 6.7: NSA: 3-step-ahead forecast for $\tilde{d}(t)$ smoothed by spline method

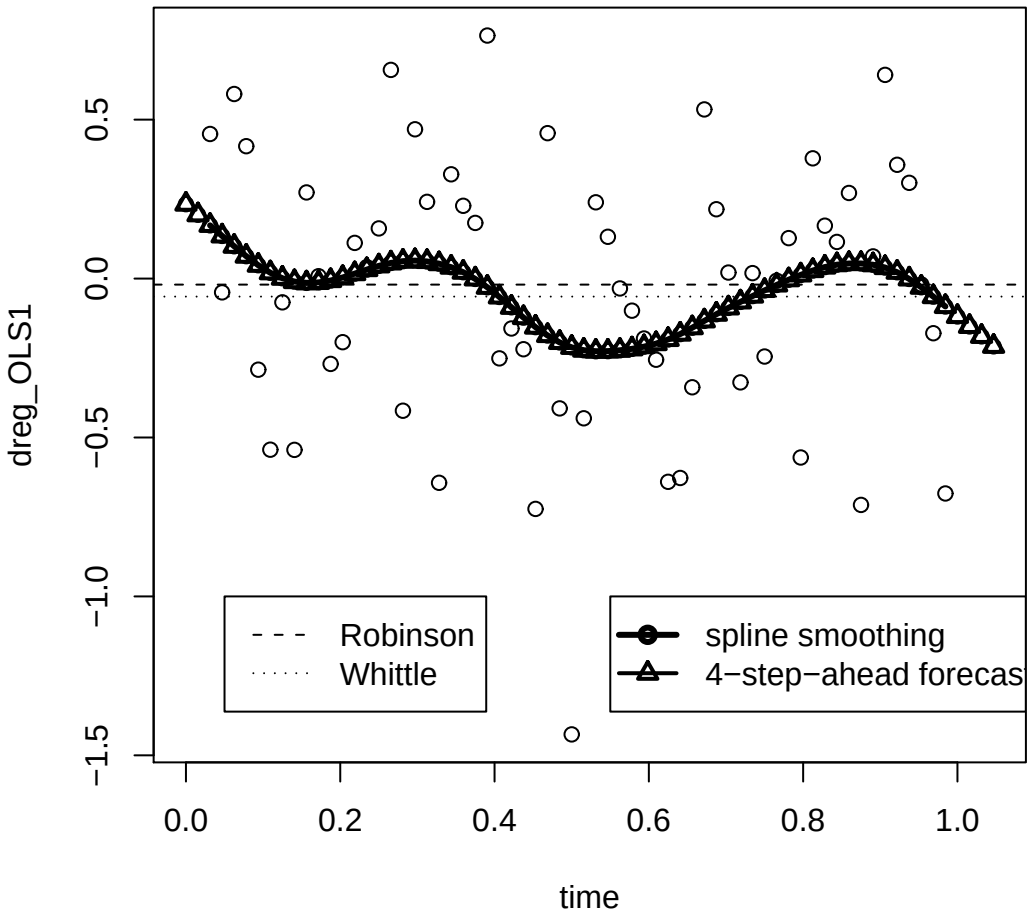


Figure 6.8: NSA: 4-step-ahead forecast for $\tilde{d}(t)$ smoothed by spline method

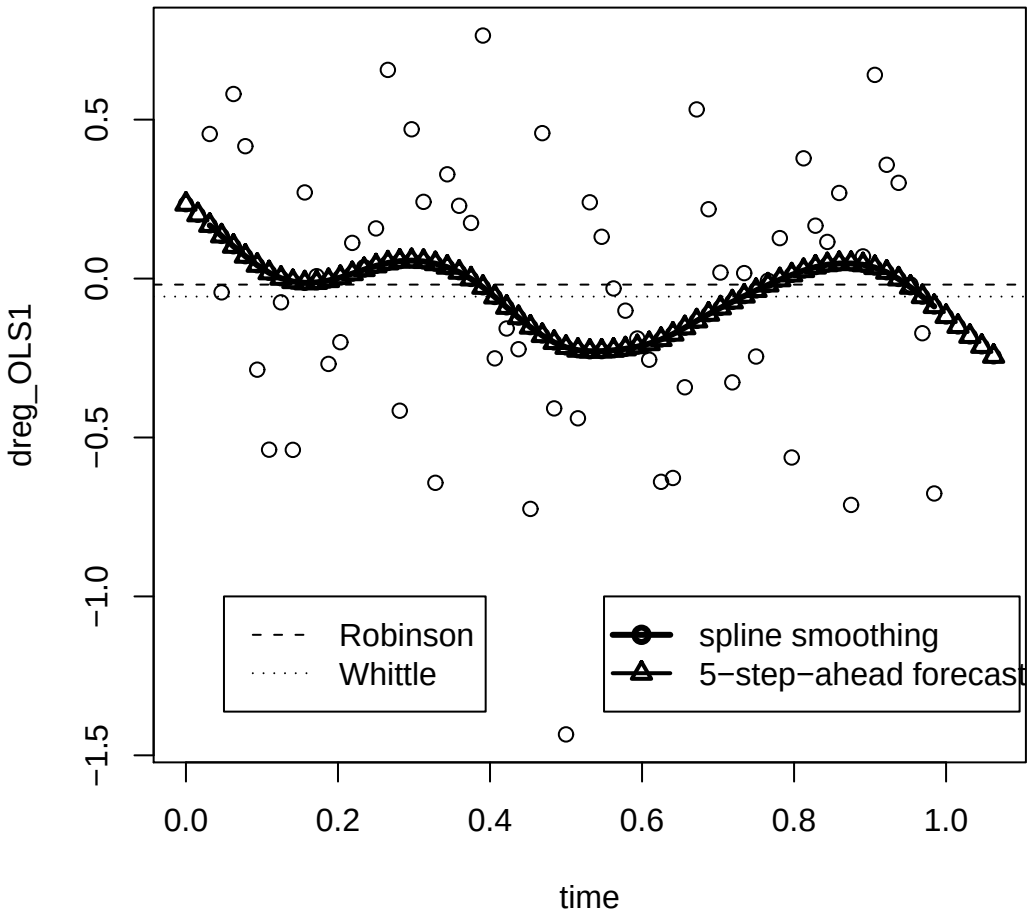


Figure 6.9: NSA: 5-step-ahead forecast for $\tilde{d}(t)$ smoothed by spline method

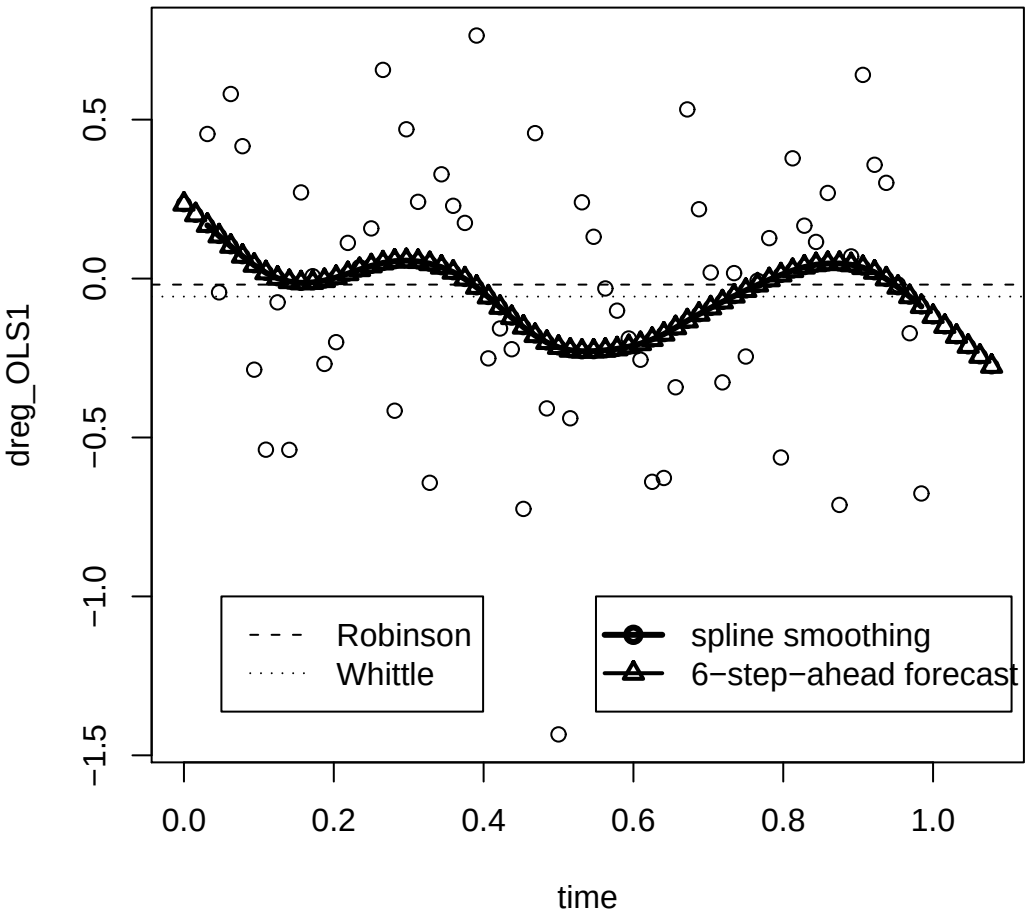


Figure 6.10: NSA: 6-step-ahead forecast for $\tilde{d}(t)$ smoothed by spline method

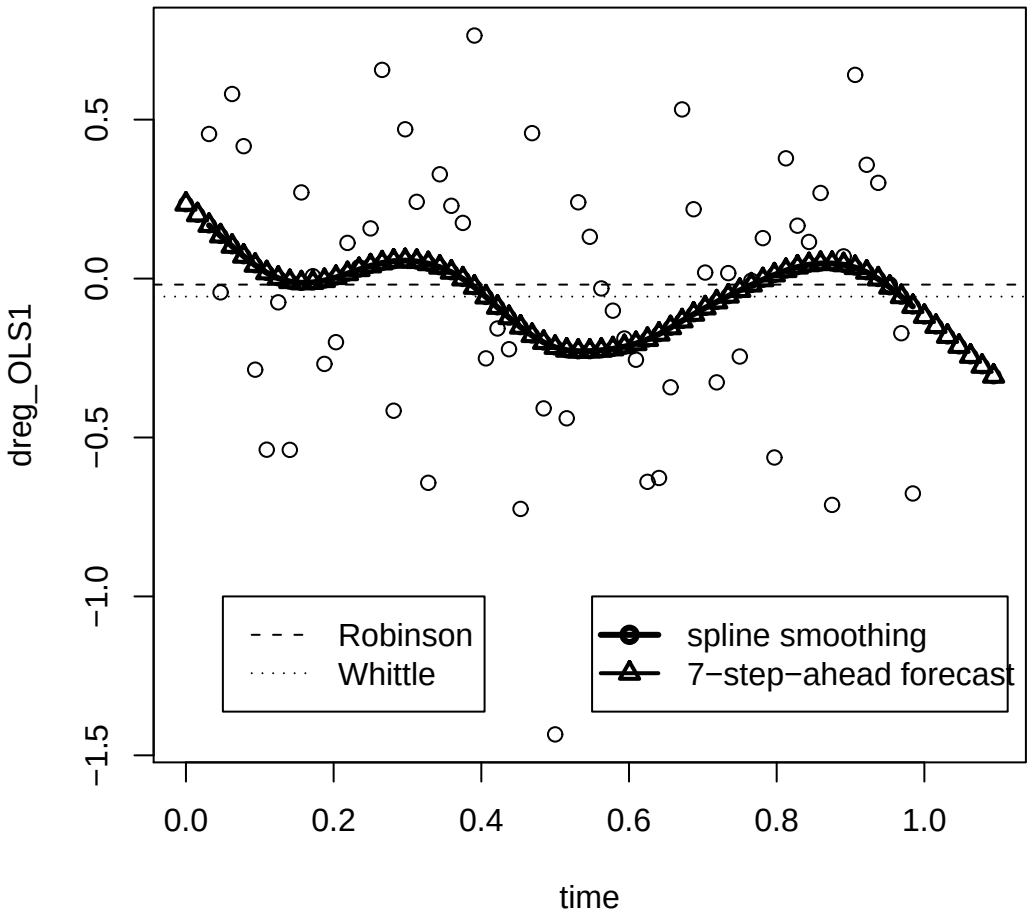


Figure 6.11: NSA: 7-step-ahead forecast for $\tilde{d}(t)$ smoothed by spline method

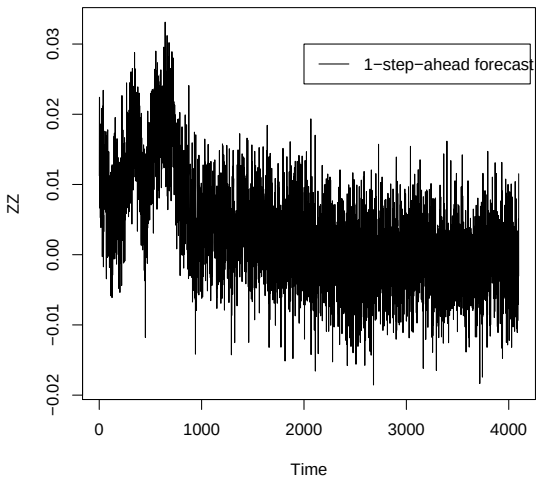


Figure 6.12: 1-step-ahead forecast for the error term in the ECM of NSA 225 index (smoothed by spline method).

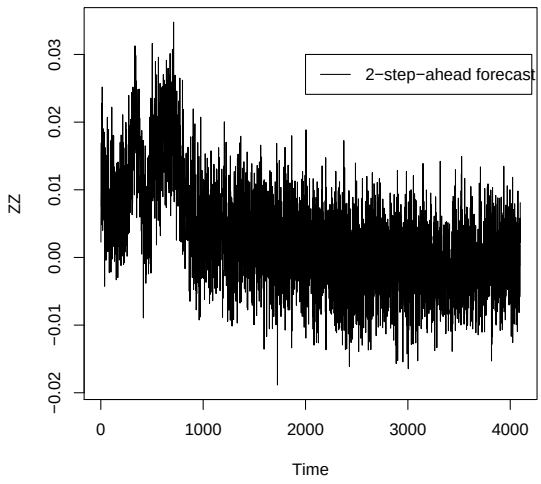


Figure 6.13: 2-step-ahead forecast for the error term in the ECM of NSA 225 index (smoothed by spline method).

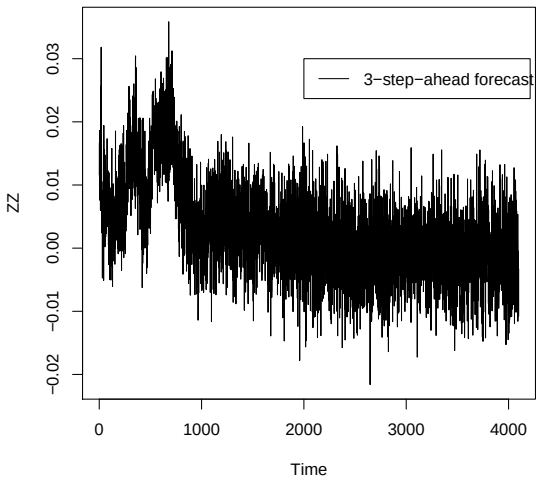


Figure 6.14: 3-step-ahead forecast for the error term in the ECM of NSA 225 index (smoothed by spline method).

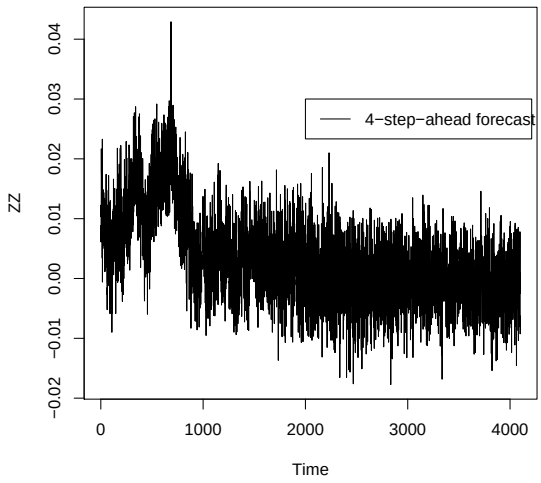


Figure 6.15: 4-step-ahead forecast for the error term in the ECM of NSA 225 index (smoothed by spline method).

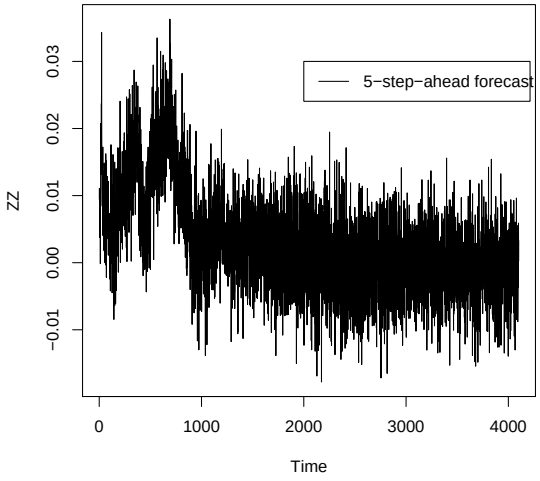


Figure 6.16: 5-step-ahead forecast for the error term in the ECM of NSA 225 index (smoothed by spline method).

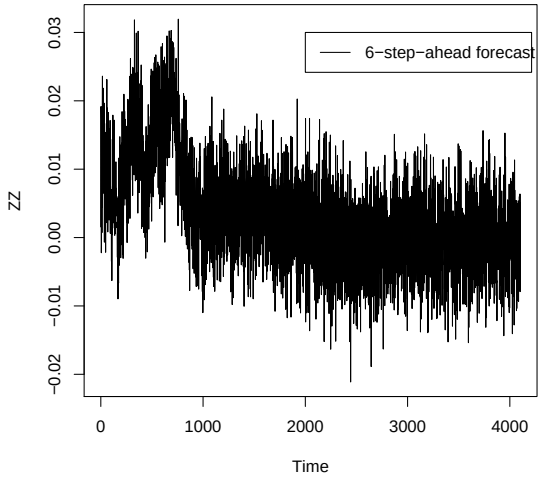


Figure 6.17: 6-step-ahead forecast for the error term in the ECM of NSA 225 index (smoothed by spline method).

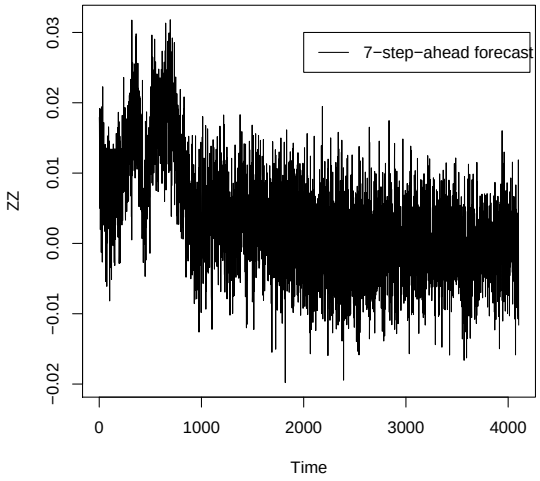


Figure 6.18: 7-step-ahead forecast for the error term in the ECM of NSA 225 index (smoothed by spline method).

	$h=1$	$h=2$	$h=3$	$h=4$	$h=5$	$h=6$	$h=7$
Bias	21.30018	21.20002	21.2284	20.79552	21.42539	21.4057	21.49394
RMSE	0.4323617	0.3029903	0.2462552	0.2110108	0.1936811	0.1756666	0.1647264

Table 6.2: The relative results of the h-step-ahead predictions on the error correction term in the ECM of NSA 225 index data using the locally stationary Gegenbauer model (parameter function smoothed by loess method).

Figure 6.28, Figure 6.29, Figure 6.30, Figure 6.31 and Figure 6.32, we can observe all the predictions.

Similarly, from the graphs, there seems to exist some differences between the true trajectory (see in Figure 6.2) and the predicted trajectory. We think that it is to some extent due to the different scales in the graphs. To have a more clear idea of our forecast, we refer to the two criteria of the forecast. In Table 6.2, the bias and the RMSE of the h-step-ahead predictions are presented. We find that the bias and the RMSE are not too large with comparison to the length of the time series Z_t, which is really satisfying.

From the numerical point of view, we find that the forecast based on the loess smoothing is really a little better than that based on the spline smoothing, which is consistent with our conjecture made from the simulation experiments for estimation.

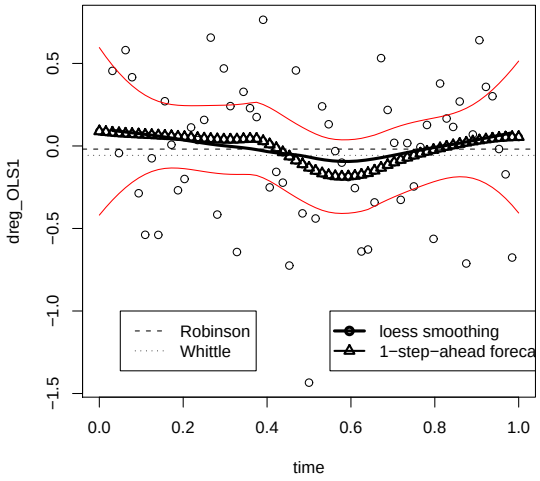


Figure 6.19: NSA: 1-step-ahead forecast for $\tilde{d}(t)$ smoothed by loess method

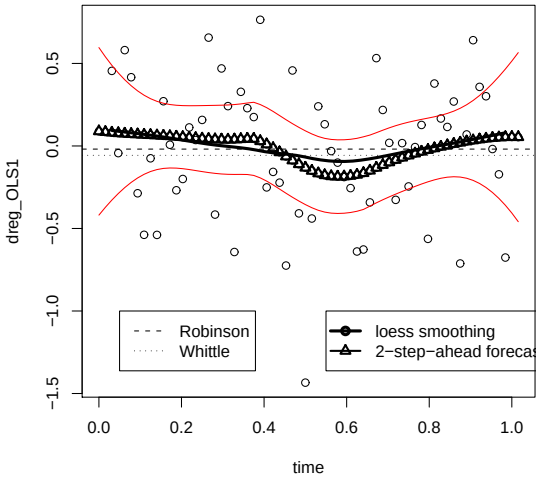


Figure 6.20: NSA: 2-step-ahead forecast for $\tilde{d}(t)$ smoothed by loess method

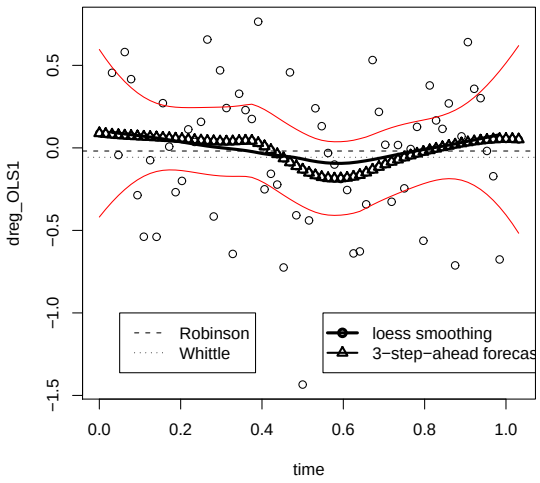


Figure 6.21: NSA: 3-step-ahead forecast for $\tilde{d}(t)$ smoothed by loess method

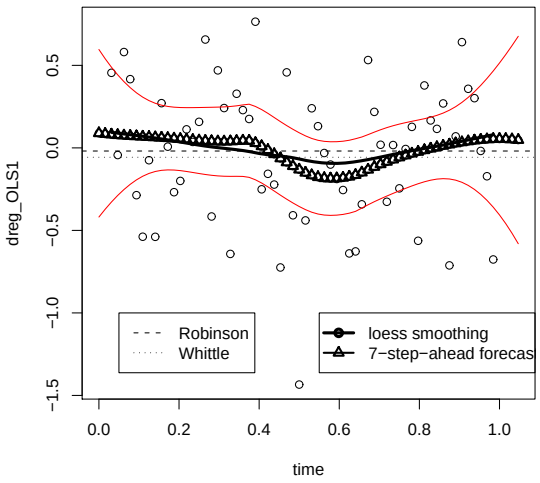


Figure 6.22: NSA: 4-step-ahead forecast for $\tilde{d}(t)$ smoothed by loess method

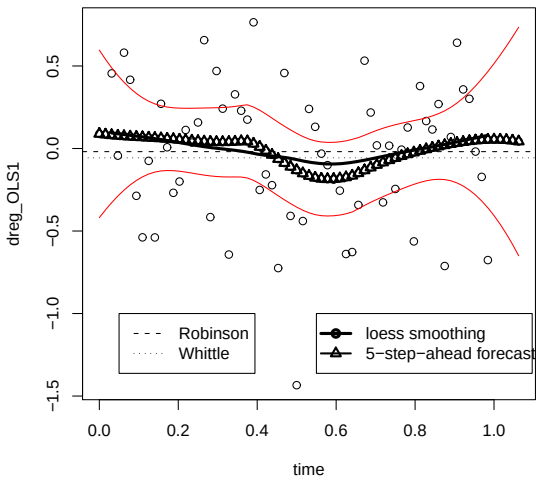


Figure 6.23: NSA: 5-step-ahead forecast for $\tilde{d}(t)$ smoothed by loess method

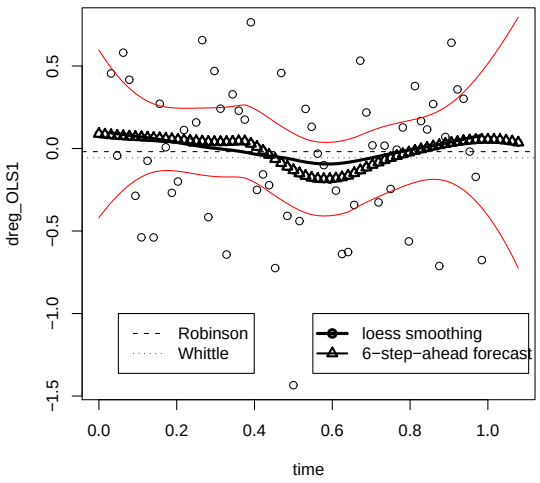


Figure 6.24: NSA: 6-step-ahead forecast for $\tilde{d}(t)$ smoothed by loess method

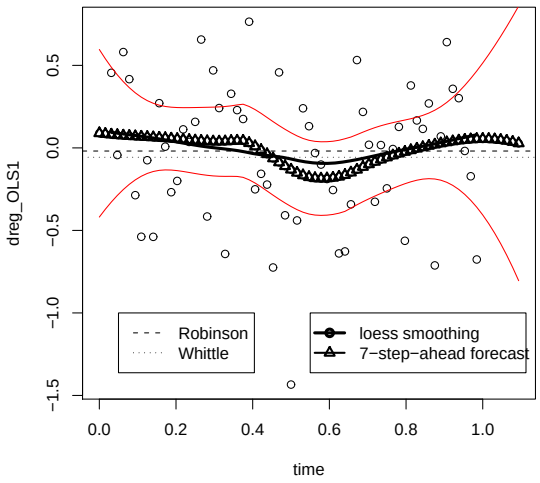


Figure 6.25: NSA: 7-step-ahead forecast for $\tilde{d}(t)$ smoothed by loess method

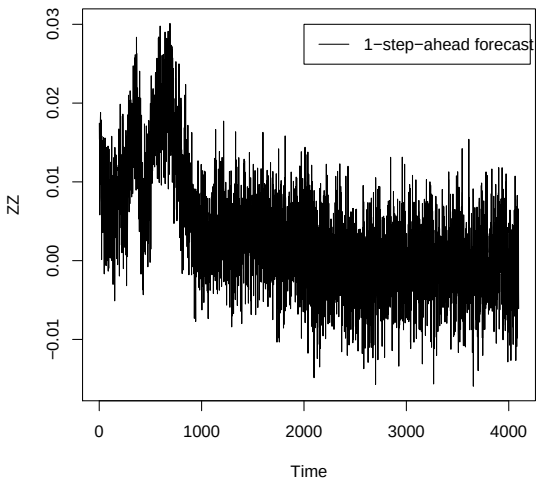


Figure 6.26: 1-step-ahead forecast for the error term in the model of NSA 225 index (smoothed by loess method).

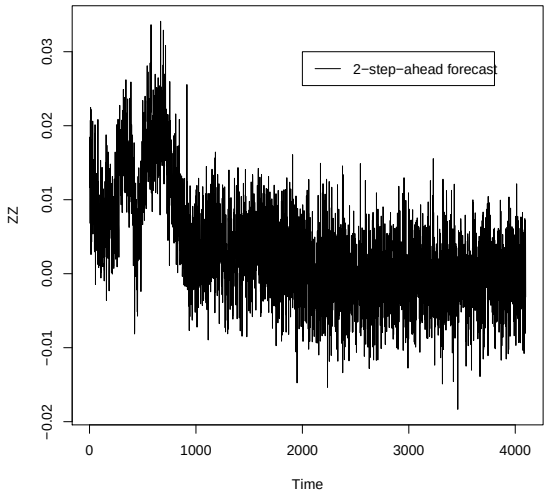


Figure 6.27: 2-step-ahead forecast for the error term in the model of NSA 225 index (smoothed by loess method).

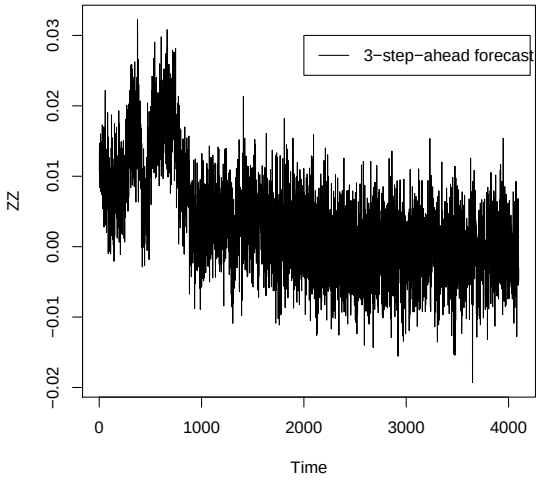


Figure 6.28: 3-step-ahead forecast for the error term in the model of NSA 225 index (smoothed by loess method).

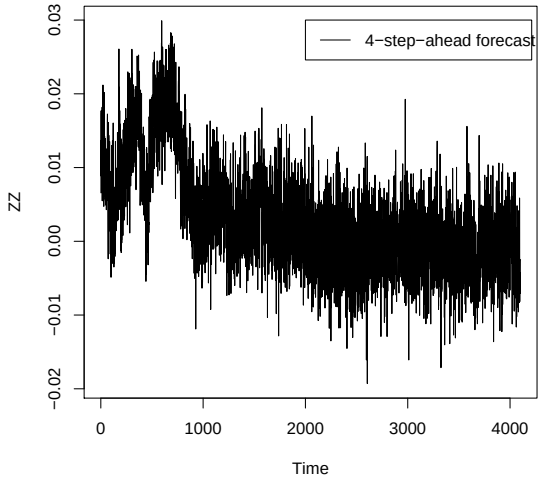


Figure 6.29: 4-step-ahead forecast for the error term in the model of NSA 225 index (smoothed by loess method).

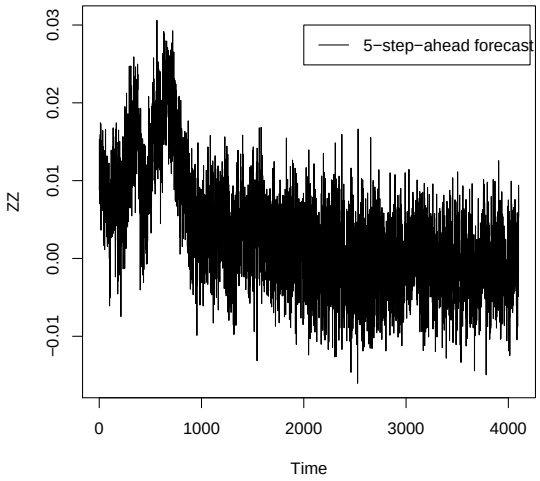


Figure 6.30: 5-step-ahead forecast for the error term in the model of NSA 225 index (smoothed by loess method).

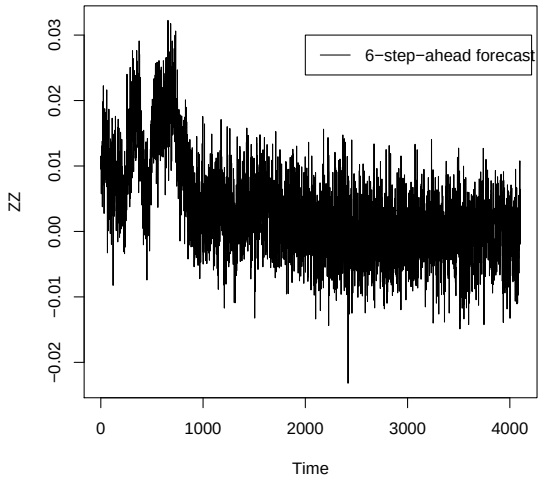


Figure 6.31: 6-step-ahead forecast for the error term in the model of NSA 225 index (smoothed by loess method).

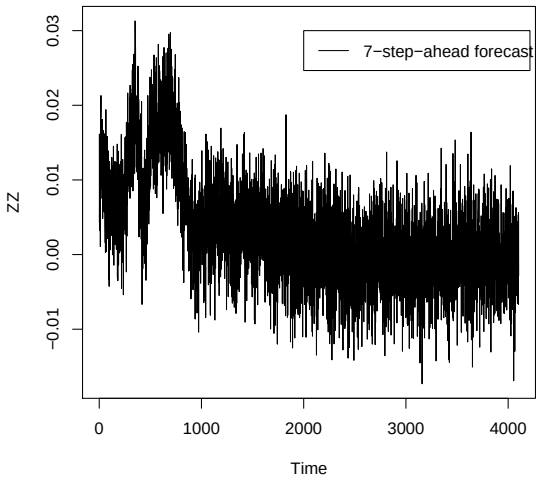


Figure 6.32: 7-step-ahead forecast for the error term in the model of NSA 225 index (smoothed by loess method).

6.2 WTI Oil Data

Crude oil prices behave similarly as many other commodities with wide price swings in times of shortage or oversupply. Thus, to predict the oil price seems to be quite important. West Texas Intermediate (WTI), also known as Texas Light Sweet, is a type of crude oil used as a benchmark in oil pricing and the underlying commodity of New York Mercantile Exchange's oil futures contracts. This oil type is often referenced in North American news reports about oil prices, alongside North Sea Brent Crude.

Historical price data for WTI can be found on a web site maintained by Energy Information Administration, Department of Energy. It is listed as WTI, Cushing, Oklahoma. Typical price difference per barrel is about $1 more than Brent, and $2 more than OPEC basket.

Denote X_t the WTI oil price. The length of the data that we consider is 5110. We observe that X_t is not stationary (see in Figure 6.33). The moments of X_t is as follows: $mean(X) = 23.76919$, $variance(X) = 100.1522$, $skewness = 1.993175$, $kurtosis = 7.363735$. We carry out the augmented Dickey-Fuller test to test the stationarity and we find that the process X_t is not stationary. Therefore, we have two choices to carry out the modeling.

First of all, we can apply the stationary models. To obtain the stationarity, we difference X_t: $U_t = X_t - X_{t-1}$. See in Figure 6.34 for more details of U_t. According to the behavior of the ACF and PACF of U_t, we decide to de-mean the series U_t: $Z_t = U_t - mean(U_t)$

and to model Z_t by the autoregressive models. The other choice is to carry out the modeling in the non-stationary setting using the new non-stationary model and new estimation algorithm that we proposed in Chapter 5.

6.2.1 Fitting by Stationary Model: AR(1)+FI(d) Model

If we chose to model Z_t by an AR(1) model, we get the following equation:

$$Z_t = \hat{a}_1 Z_{t-1} + \varepsilon_t,$$

where $\hat{a}_1 = -0.04144$. In Figure 6.35, we can find how the AR(1) model fits Z_t. According to the fitting behavior of Z_t, we turn to study the residuals $(\varepsilon_t)_t$. The related information for ε_t is exhibited in Figure 6.36. It is something like a noise. Thus we study the square of ε_t to investigate the properties of the volatility in the oil price, see in Figure 6.37 for the information of $(\varepsilon_t^2)_t$. Since there exists an explosion near the 0 frequency in its spectrum, we decide to fit the volatility $(\varepsilon_t^2)_t$ by a fractional integrated model FI(d) as follows:

$$(I - B)^{\hat{d}} \varepsilon_t^2 = \nu_t,$$

where $\hat{d} = 0.1234373$. Observing Figure 6.38, we do not find the existence of long memory behavior in the residual ν_t. However, ν_t is not totally white noise. Since the residual ν_t is a a little complicated, here we just consider it as a Gaussian white noise for simplicity.

Thus, we model Z_t as follows:

$$\begin{cases} Z_t = \hat{a}_1 Z_{t-1} + \varepsilon_t \\ (I - B)^{\hat{d}} \varepsilon_t^2 = \nu_t, \end{cases} \tag{6.3}$$

where $\hat{a}_1 = -0.0414$, $\hat{d} = 0.1234373$ and ν_t is Gaussian white noise with mean zero.

Graphically, it seems that this model does not fit very well the series $(Z_t)_t$.

6.2.2 Fitting by Stationary Model: AR(2)+FI(d) Model

If we model Z_t by an AR(2) model, we get the following equation:

$$Z_t = \hat{a}_1 Z_{t-1} + \hat{a}_2 Z_{t-2} + \varepsilon_t,$$

where $\hat{a}_1 = -0.0422963$, $\hat{a}_2 = -0.02060408$. In Figure 6.39, we can observe how Z_t is fit by the AR(2) model. According to the fitting behavior of Z_t, we turn to study the residuals $(\varepsilon_t)_t$. The related information of $(\varepsilon_t)_t$ is exhibited in Figure 6.40. It looks something like a noise,thus we study the square of the $(\varepsilon_t)_t$ to investigate the properties of the volatility in the oil price, see in Figure 6.41 for the information of $(\varepsilon_t^2)_t$. We observe that there exists an explosion near the 0 frequency in the spectrum, thus we fit the volatility $(\varepsilon_t^2)_t$ by a fractional integrated model FI(d) as follows:

$$(I - B)^{\hat{d}} \varepsilon_t^2 = \nu_t,$$

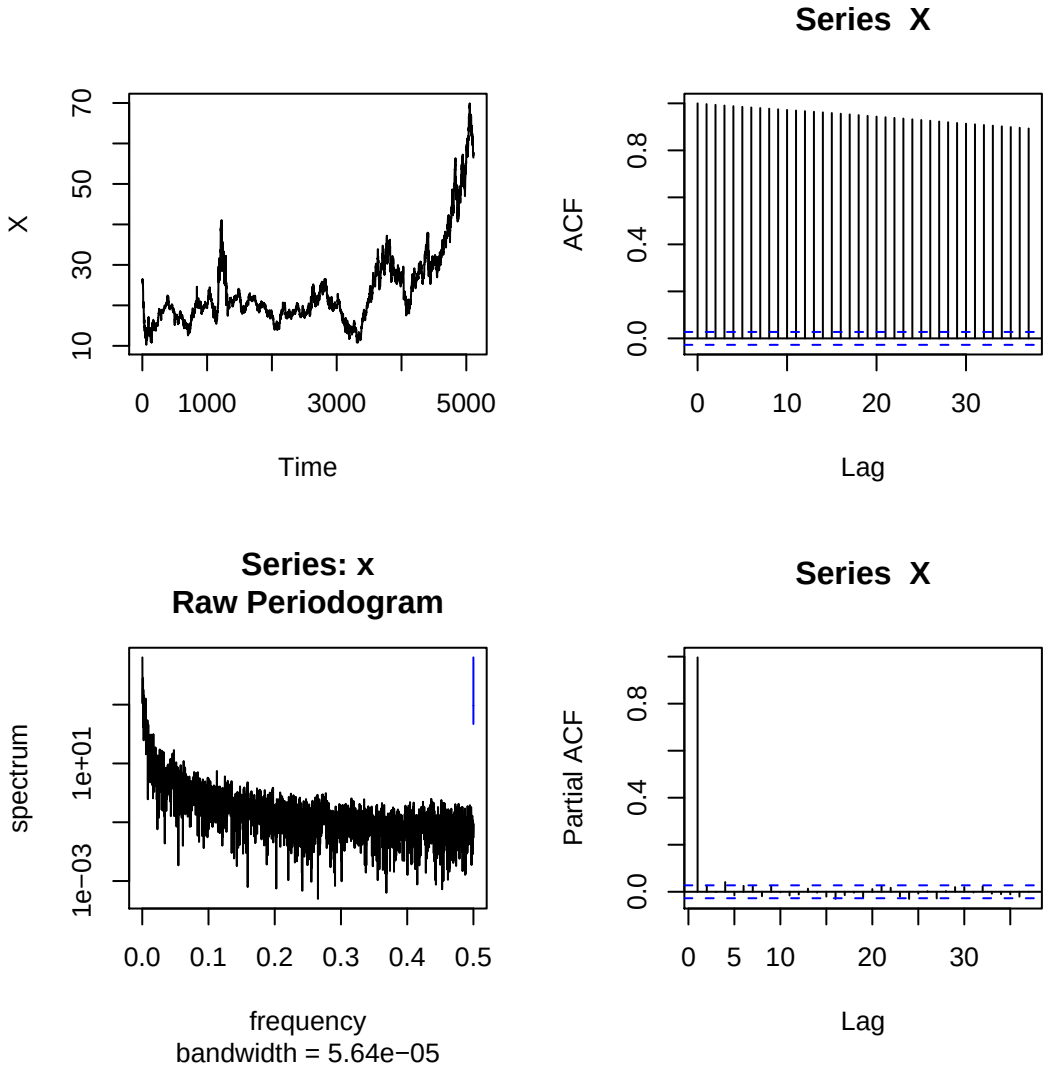


Figure 6.33: WTI: The sample path, spectrum, ACF and PACF of $(X_t)_t$

where $\hat{d} = 0.1230429$, estimated by the GPH method. In the residual ν_t, there does not exist long memory behavior, see Figure 6.42. We can regard it as a kind of noise, but it is not completely a white noise. For simplicity, we just consider it as a Gaussian white noise.

Thus, we model Z_t as follows:

$$\begin{cases} Z_t = \hat{a}_1 Z_{t-1} + \hat{a}_2 Z_{t-2} + \varepsilon_t \\ (I - B)^{\hat{d}} \varepsilon_t^2 = \nu_t, \end{cases} \tag{6.4}$$

where $\hat{a}_1 = -0.0422963$, $\hat{a}_2 = -0.02060408$, $\hat{d} = 0.1230429$ and ν_t is Gaussian white noise with zero mean.

Similarly, from the graphs, it seems that this model does not fit very well the series Z_t.

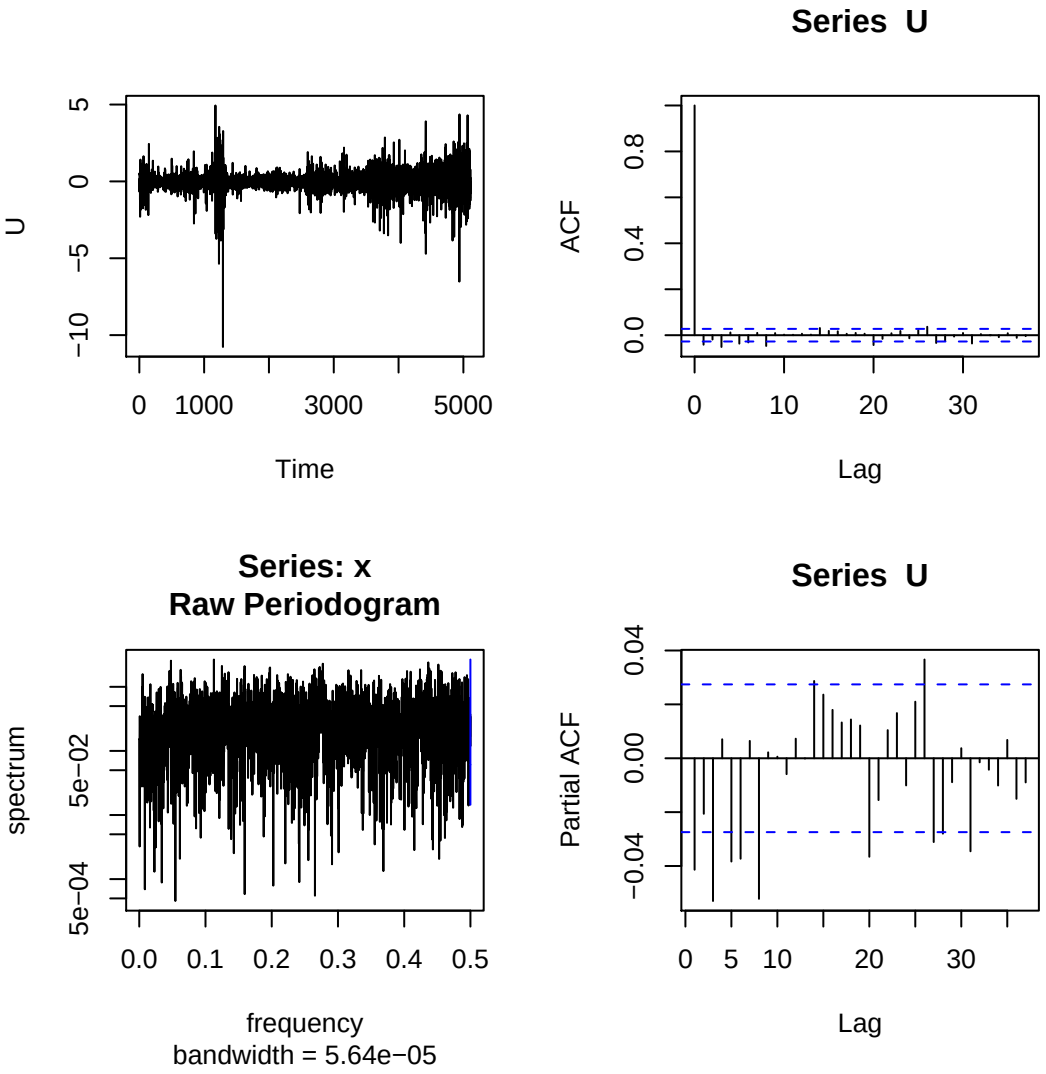


Figure 6.34: WTI: The sample path, spectrum, ACF and PACF of $(U_t)_t$

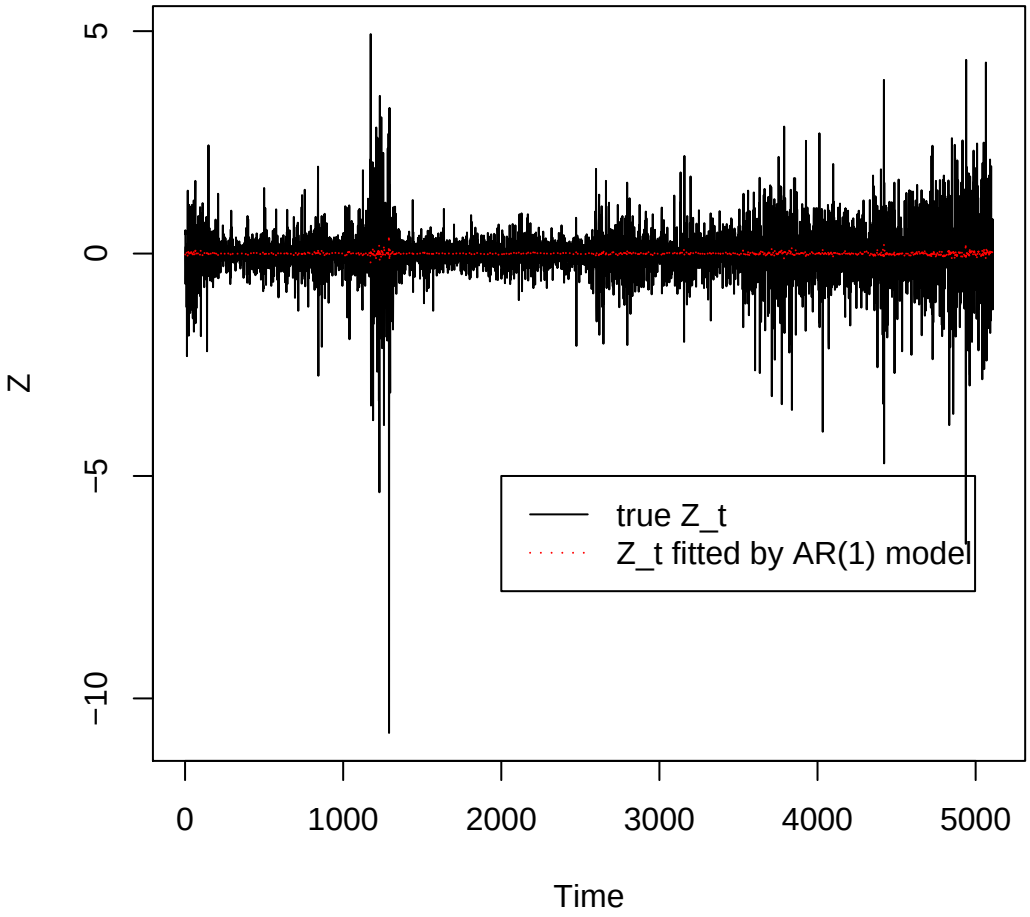


Figure 6.35: WTI: Fit Z_t by AR(1) model

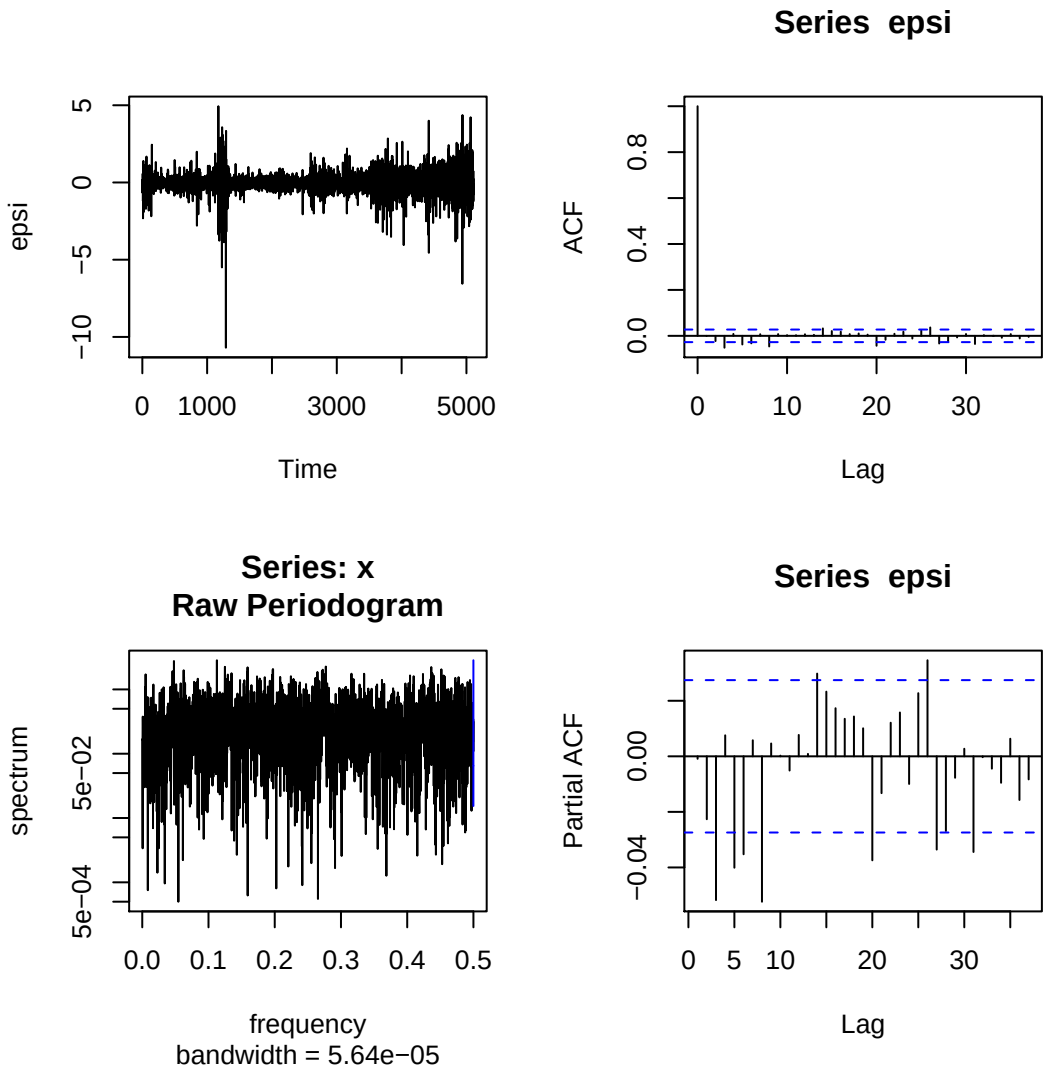


Figure 6.36: WTI: AR(1)+FI(d), The sample path, spectrum, ACF and PACF of the residual $(\varepsilon_t)_t$ of the AR(1) term

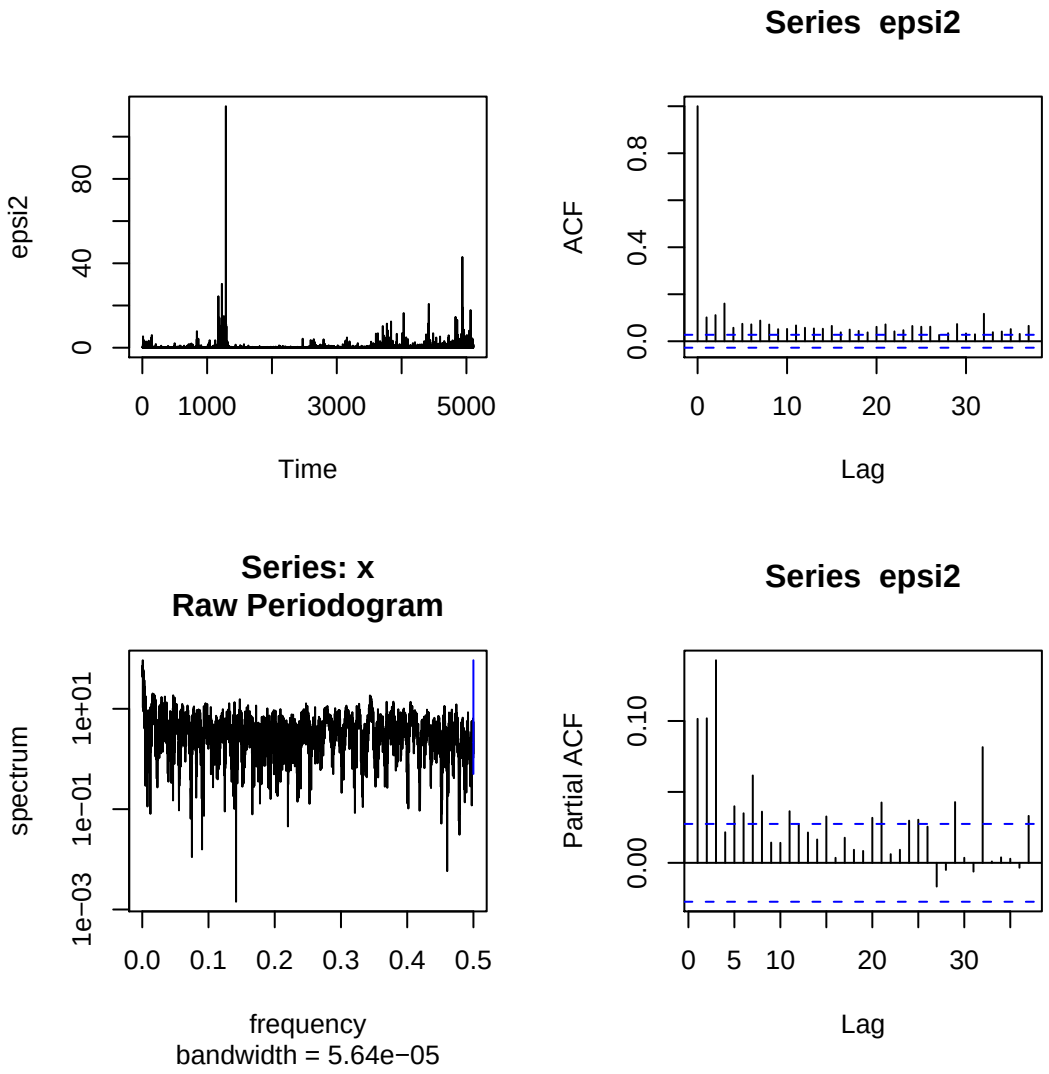


Figure 6.37: WTI: AR(1)+FI(d), The sample path, spectrum, ACF and PACF of the volatility $(\varepsilon_t^2)_t$

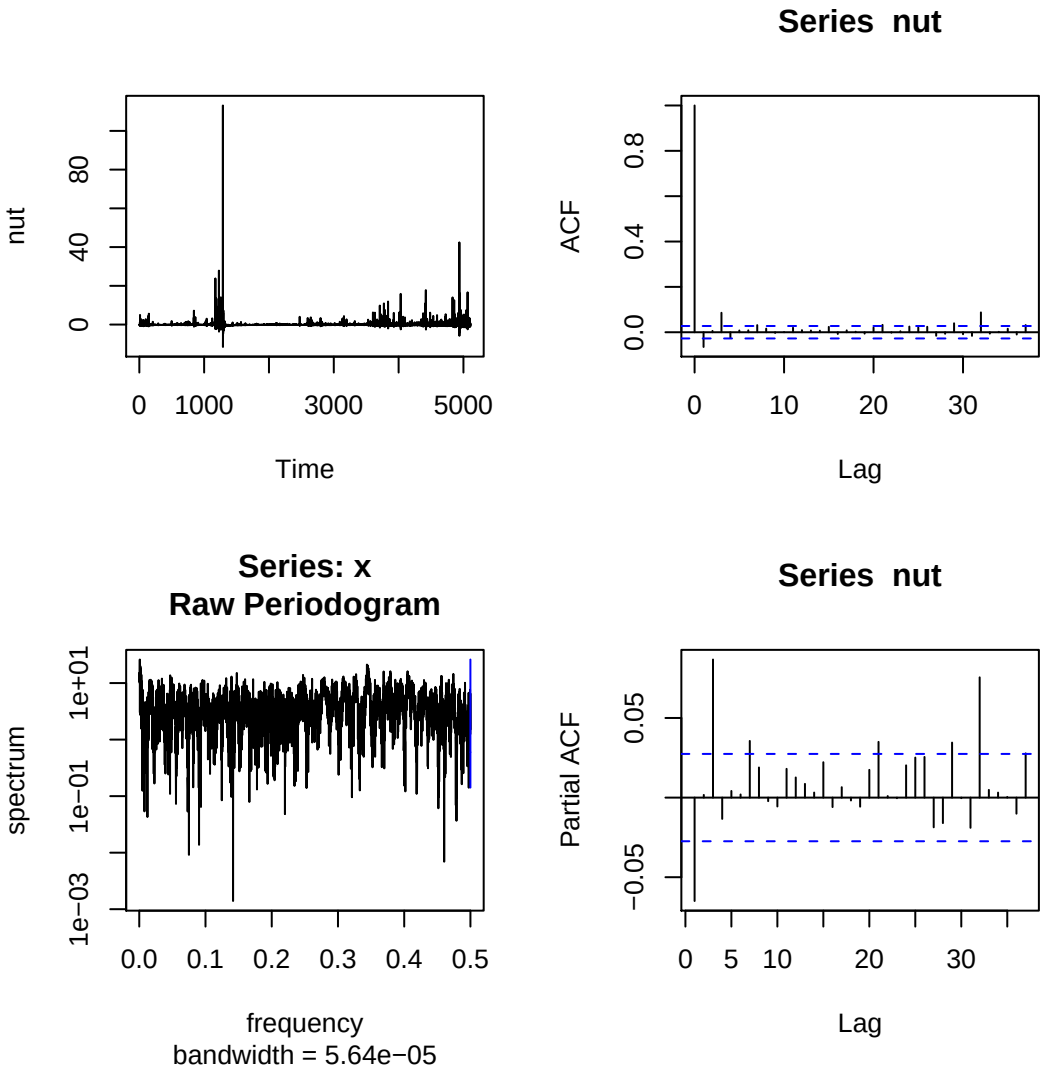


Figure 6.38: WTI: AR(1)+FI(d), The residual $(\nu_t)_t$ of the FI(d) term

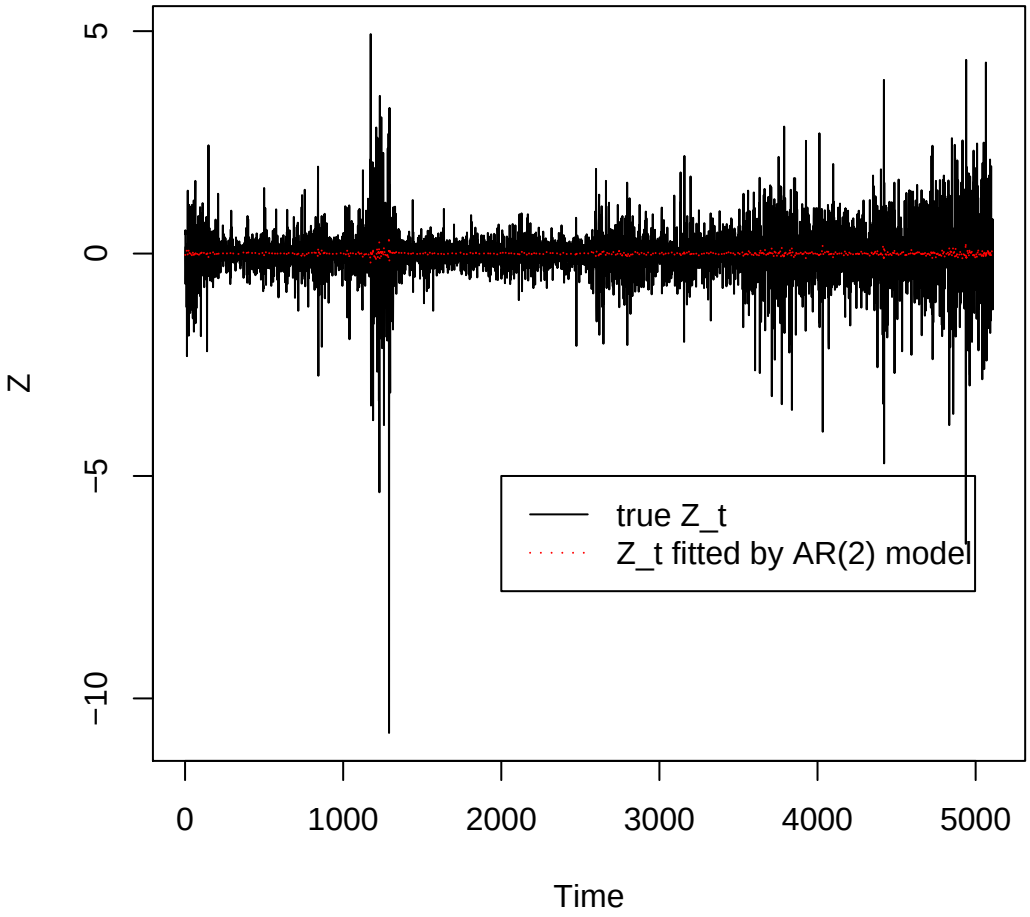


Figure 6.39: WTI: Fit Z_t by AR(2) model

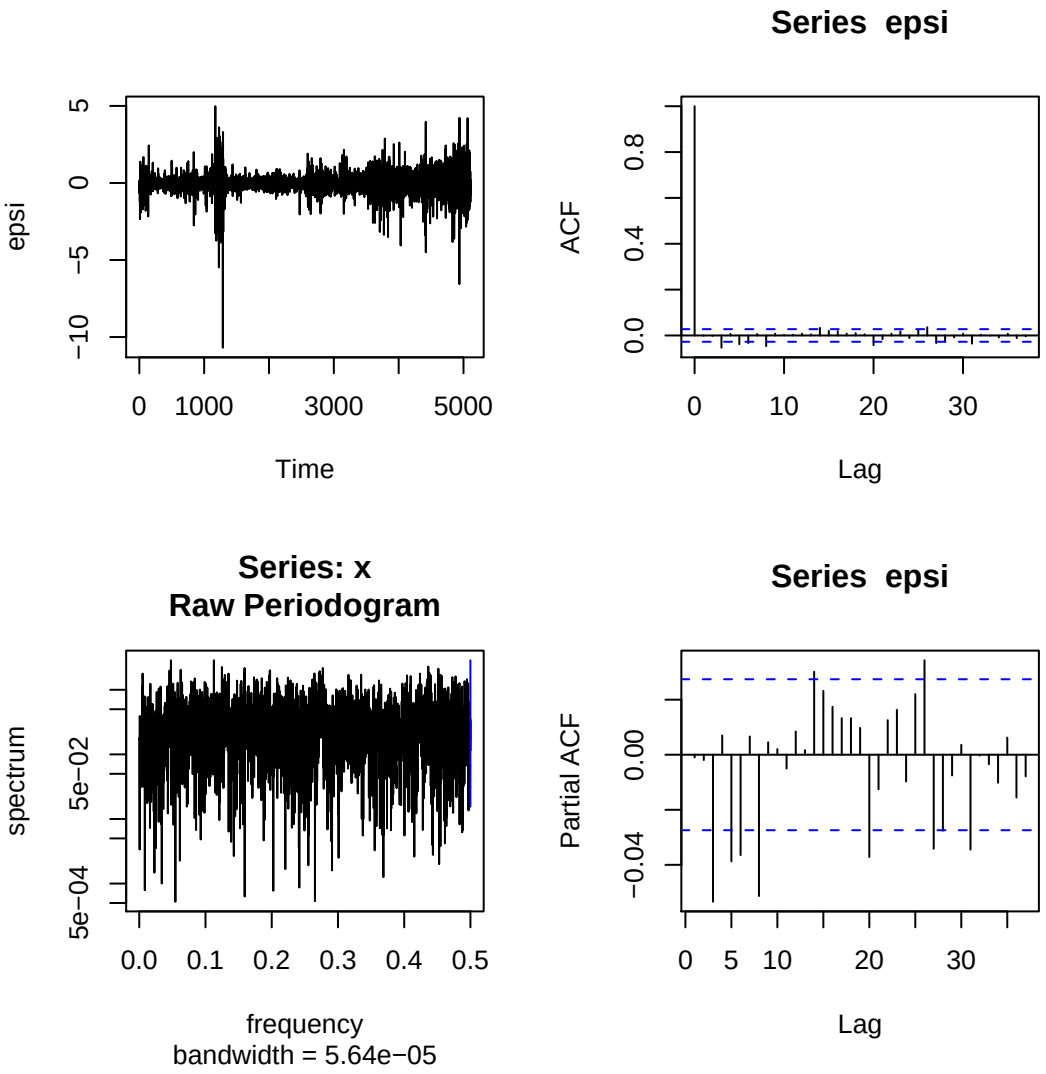


Figure 6.40: WTI: AR(2)+FI(d), Sample path, spectrum, ACF and PACF of the residual $(\varepsilon_t)_t$ of AR(2) term

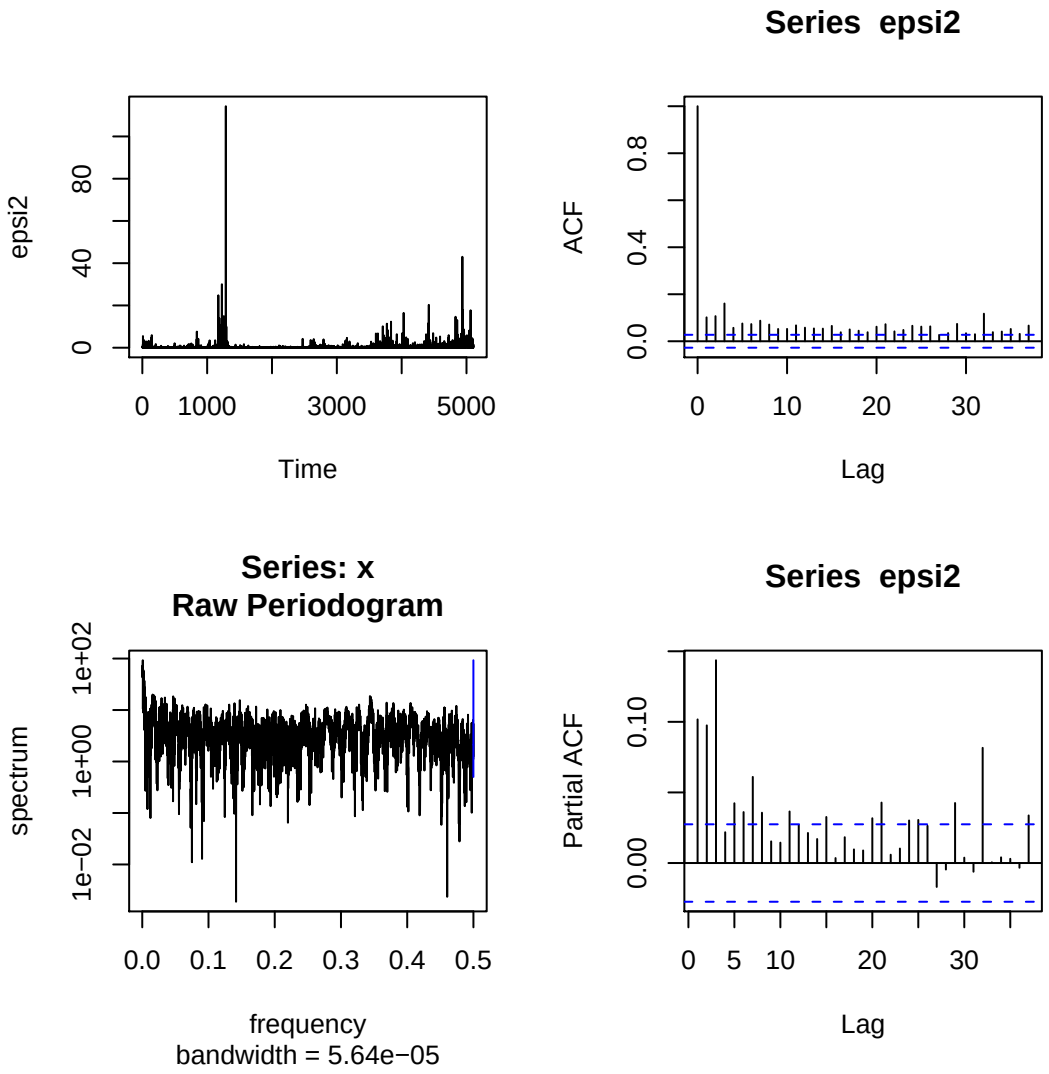


Figure 6.41: WTI: AR(2)+FI(d), The sample path, spectrum, ACF and PACF of the volatility $(\varepsilon_t^2)_t$

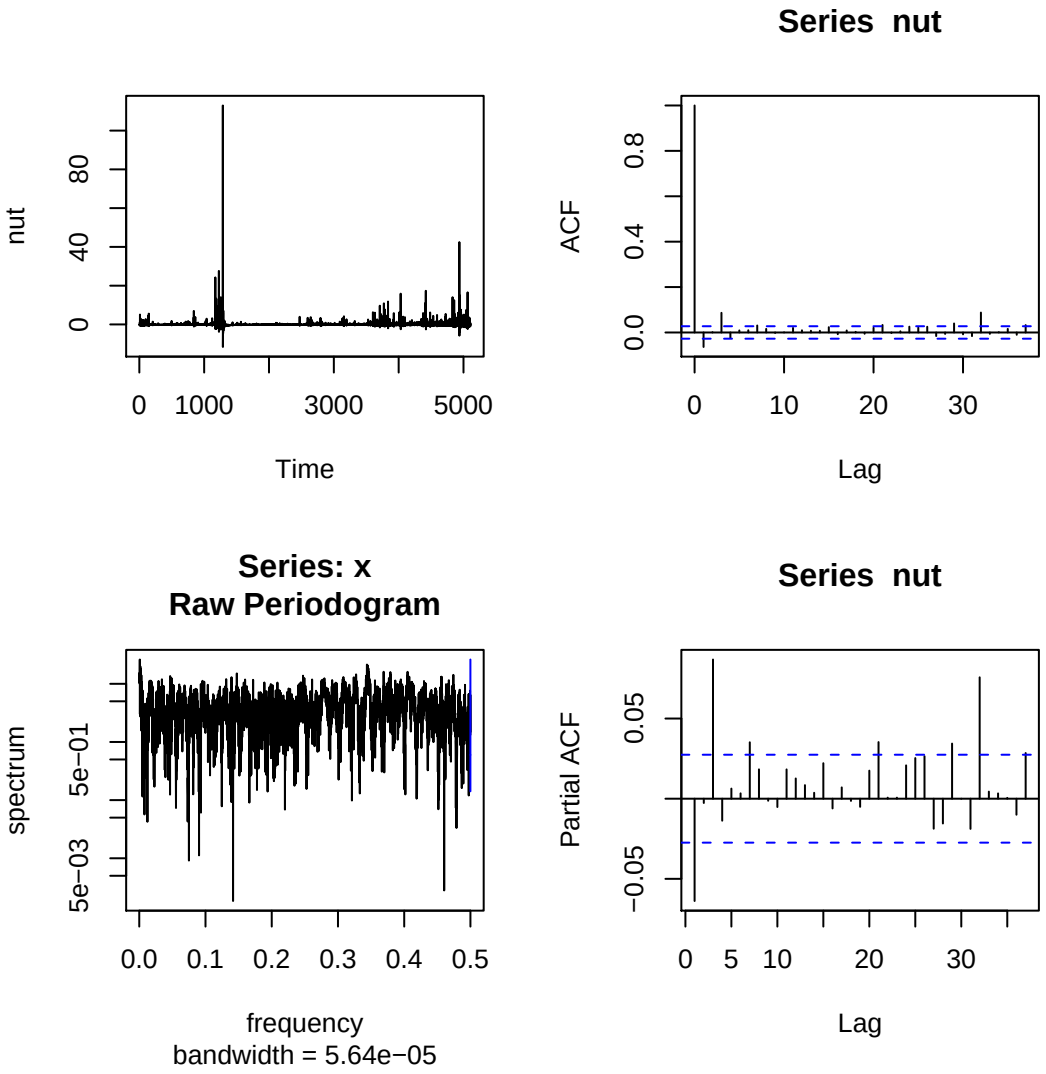


Figure 6.42: WTI: AR(2)+FI(d), The residuals of the FI(d) term $(\nu_t)_t$

6.2.3 Fitting by Non-stationary Model Using Wavelet Method

First we detrend X_t through multiresolution analysis (MRA) by applying the MODWT on X_t, where we adopt the Daubechies least asymmetric "LA8" as the wavelet filter, the depth of the decomposition is 6 and the decomposed vector is assumed to be periodic on its defined interval. That is to say, $Z_t = X_t - S_{6,t}$, where $S_{6,t}$ is the wavelet smooth in the multiresolution analysis presented in Figure 6.43. We investigate the properties of Z_t in Figure 6.44. Observing the behavior of ACF and spectrum of Z_t, we decide to model Z_t by locally stationary 1-factor Gegenbauer process with time-varying parameter function:

$$(I - 2\nu B + B^2)^{d(t)} Z_t = \varepsilon_t.$$

We estimate the Gegenbauer frequency which corresponds to the frequency of the maximum periodogram and we get $\hat{\lambda}_G = 0.01049805$. Then we apply the wavelet-based algorithm that we proposed in Chapter 5. We partition evenly the time interval and carry out the OLS regression using DWPT wavelet coefficients on each subinterval and we smooth the estimations by spline method and loess method.

Thus we model X_t as follows:

$$\begin{cases} X_t = Z_t + S_{6,t} \\ (I - 2\nu B + B^2)^{\hat{d}(t)} Z_t = \varepsilon_t, \end{cases} \tag{6.5}$$

where $\nu = \cos(2\pi\hat{\lambda}_G) = 0.9978253$, ε_t is Gaussian white noise with zero mean and the estimation $\hat{d}(t)$ is presented in the Figure 6.45. The thin curve is the estimation smoothed by spline method. The thick curve is the estimation smoothed by loess method.

If we observe carefully the Figure 6.45, we find that the estimated function $\hat{d}(t)$ does not lie well in the definition interval $(-\frac{1}{2}, \frac{1}{2})$ for both two kinds of smoothing method. It is really a little awkward. To solve this problem, we decide to do some differencing in order to satisfy the range of definition interval for the parameter function. After differencing, we investigate the properties of Z_t in Figure 6.46. Thus we have the modeling as follows:

$$\begin{cases} X_t = Z_t + S_{6,t} \\ (I - 2\nu B + B^2)^{\hat{d}(t)+1} Z_t = \varepsilon_t, \end{cases} \tag{6.6}$$

where $\nu = \cos(2\pi\hat{\lambda}_G) = 0.9978253$, ε_t is Gaussian white noise with zero mean and the estimation $\hat{d}(t)$ is presented in the Figure 6.47. The thin curve is the estimation smoothed by spline method. The thick curve is the estimation smoothed by loess method.

Now the estimated parameter function smoothed by the loess method lies well in the definition interval, while part of the parameter function smoothed by the spline is still outside the interval $(-\frac{1}{2}, \frac{1}{2})$.

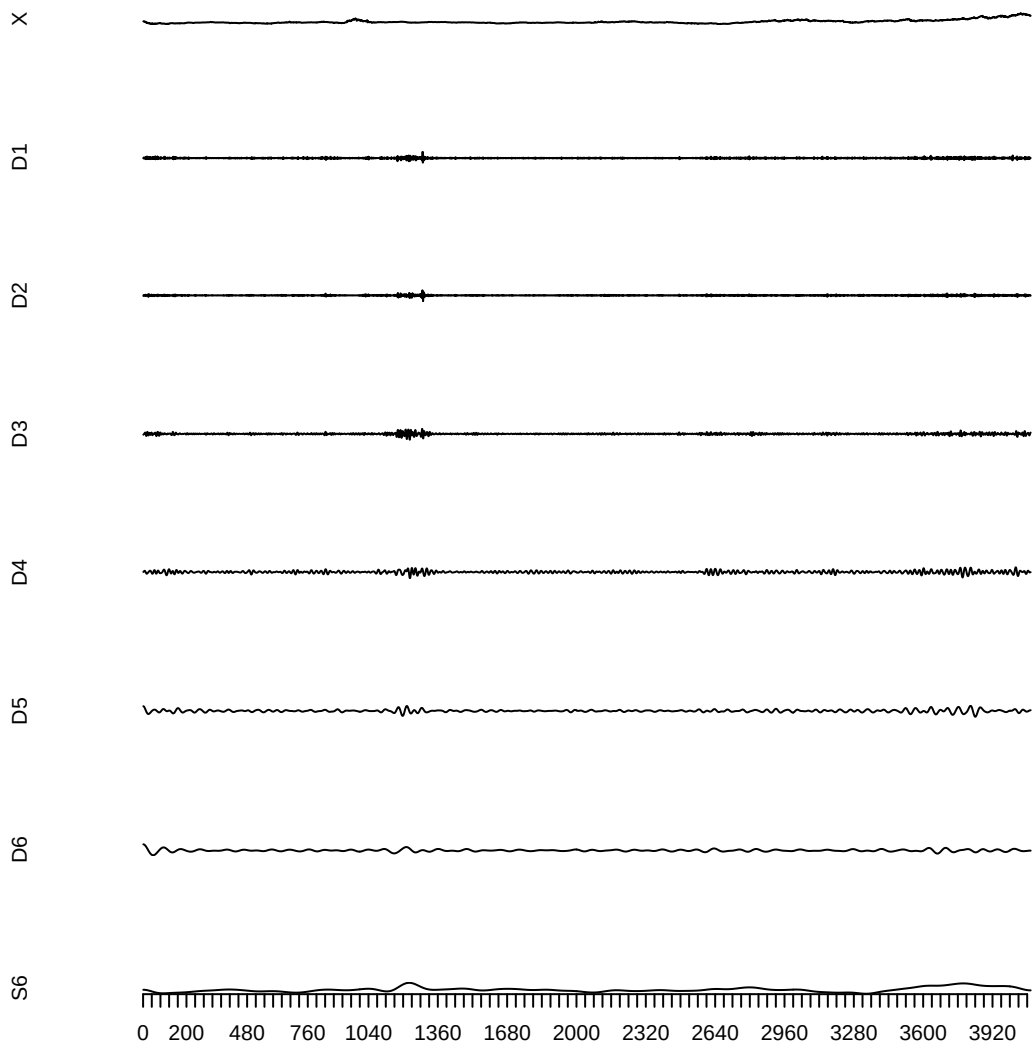


Figure 6.43: WTI: The Multiresolution analysis of $(X_t)_t$

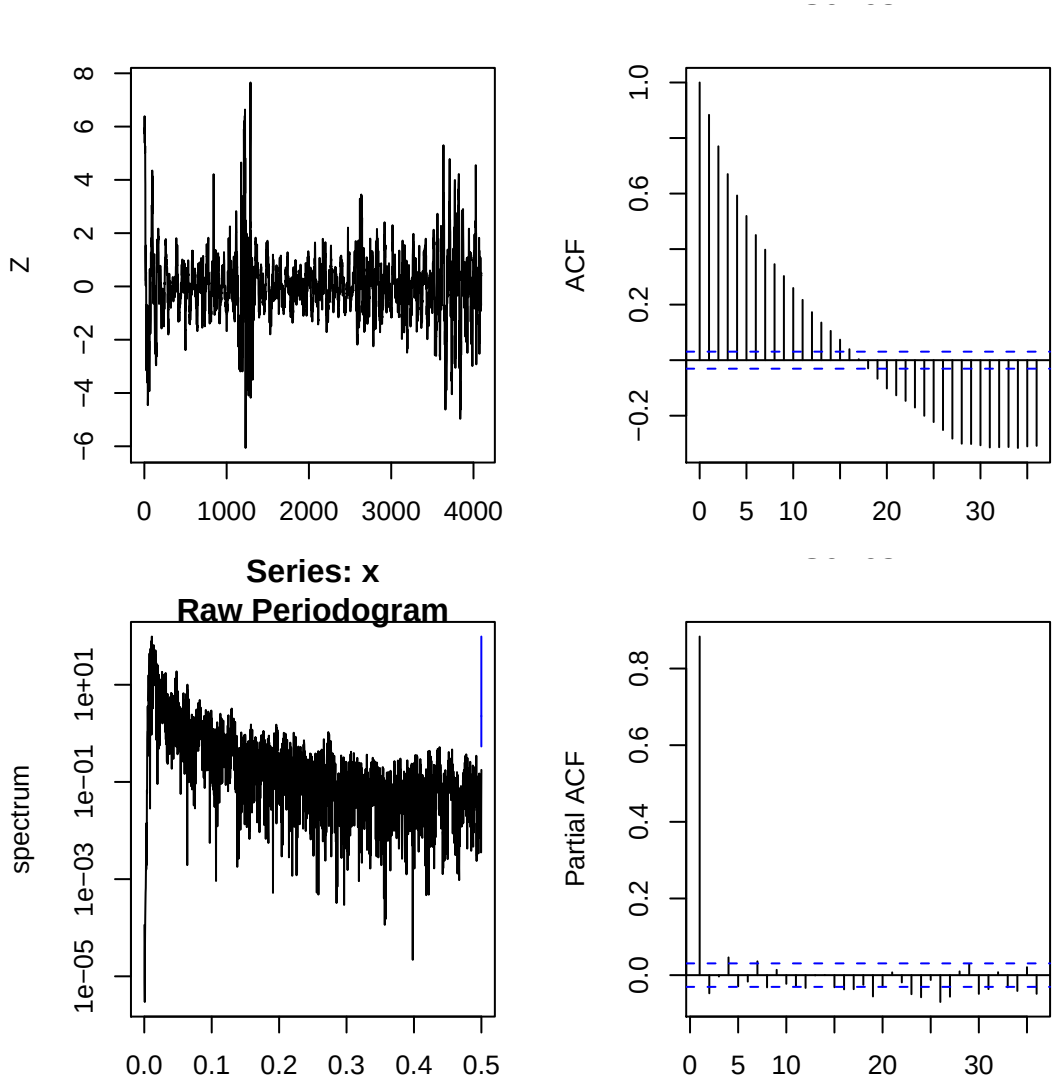


Figure 6.44: WTI: The sample path, spectrum, ACF and PACF of $(Z_t)_t$

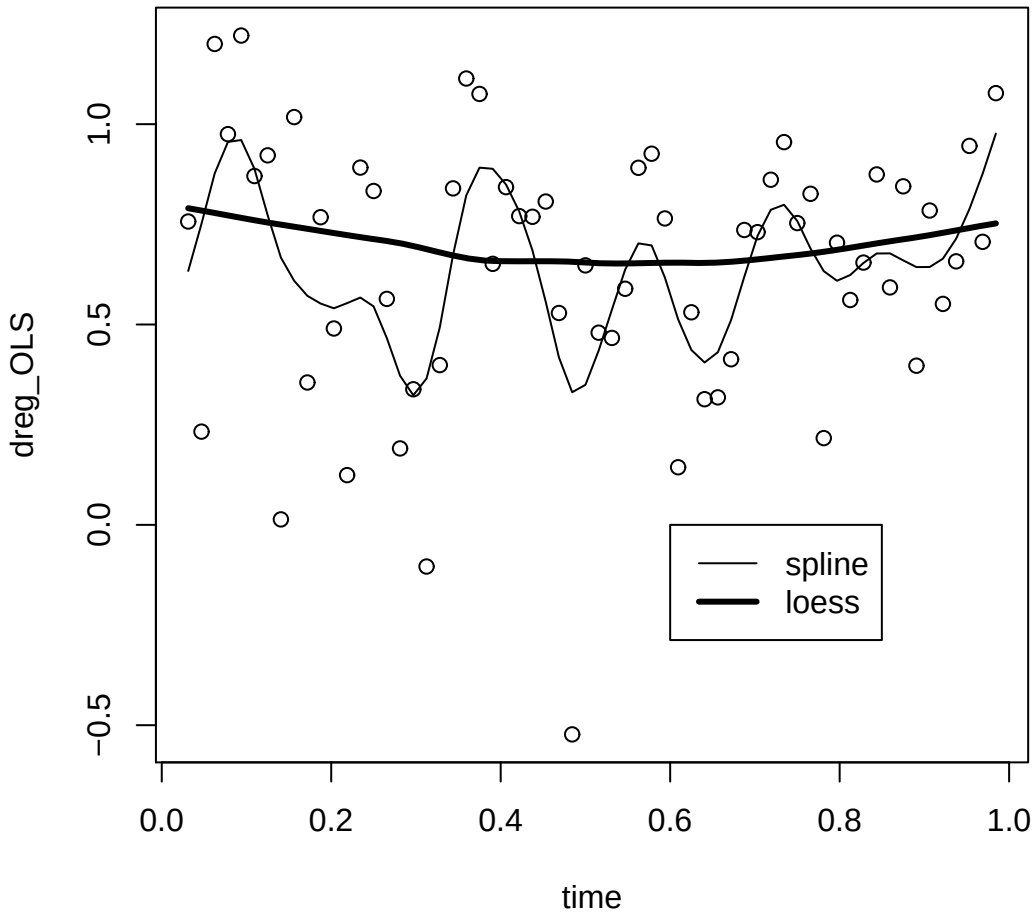


Figure 6.45: WTI: The estimated parameter of $(d_t)_t$

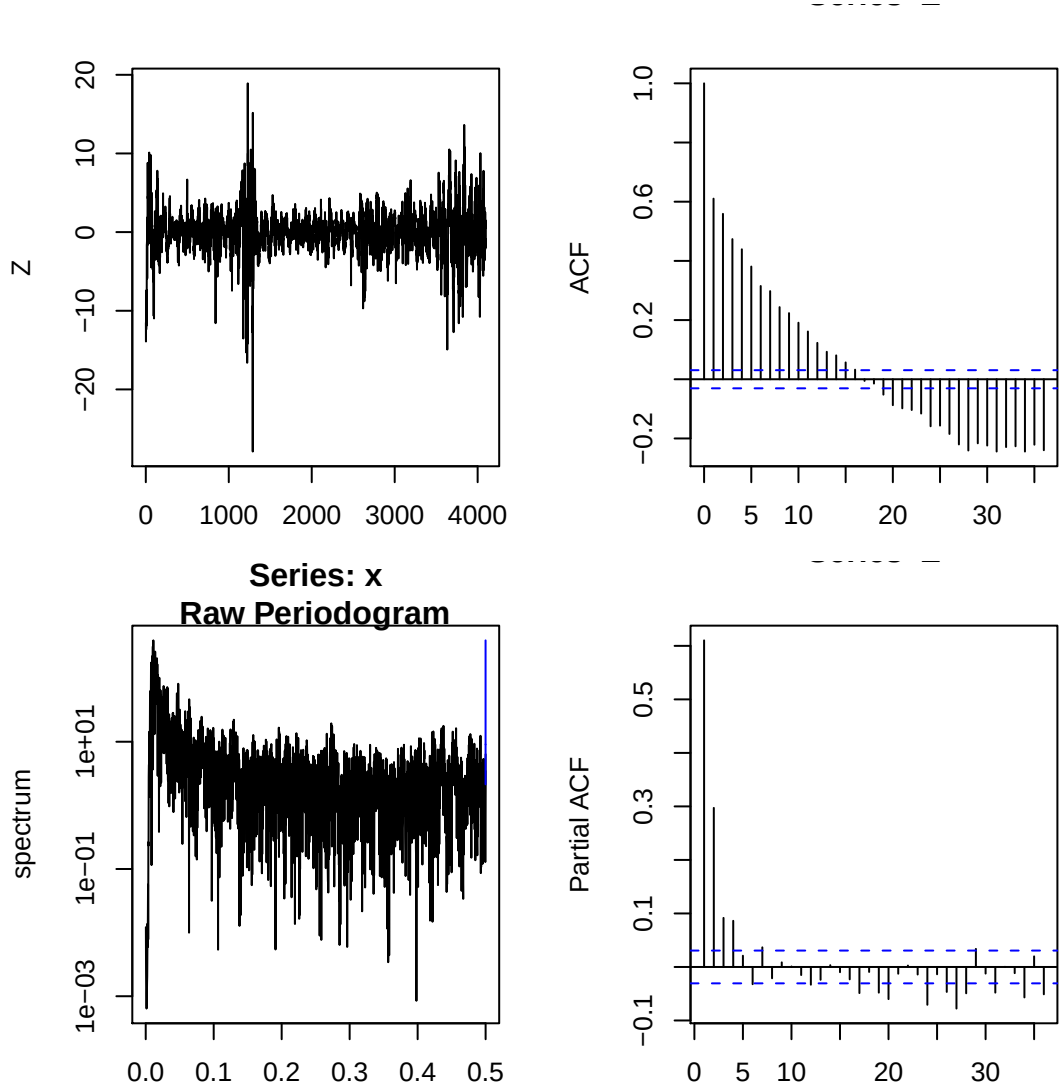


Figure 6.46: WTI: The sample path, spectrum, ACF and PACF of $(Z_t)_t$ after differencing

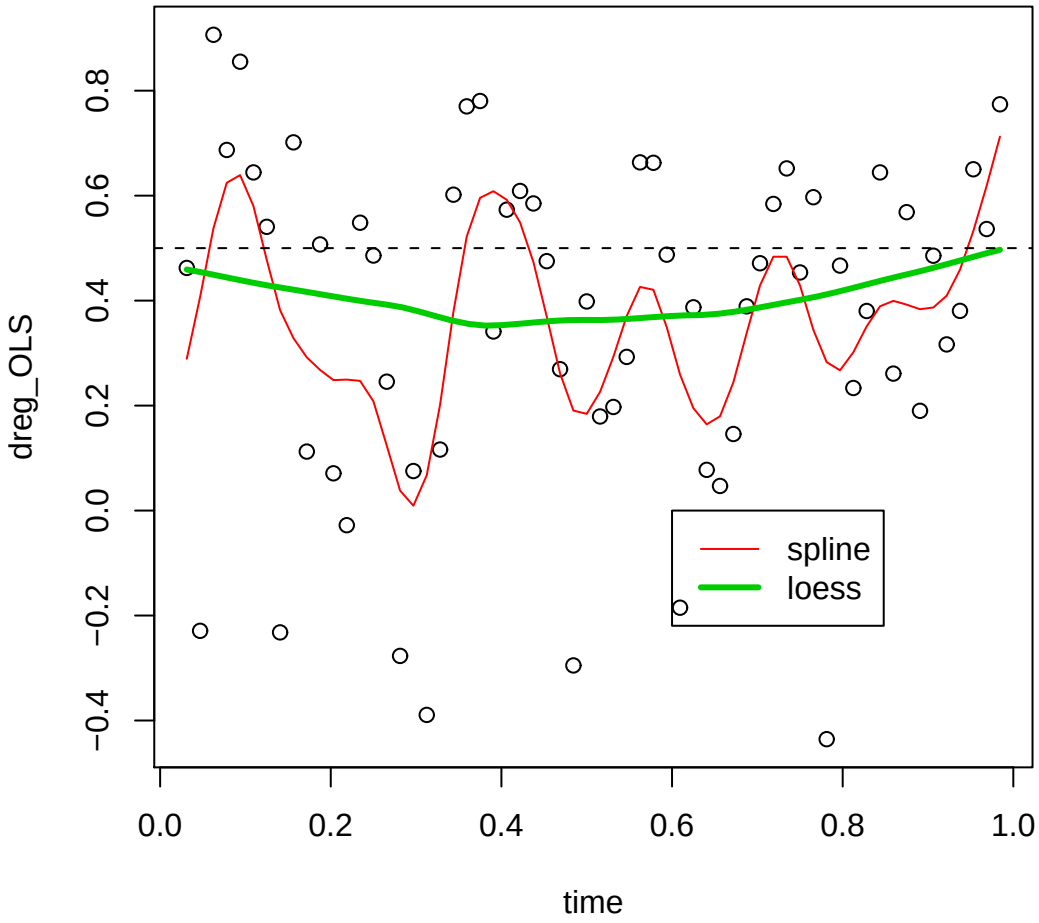


Figure 6.47: WTI: The estimated parameter of $(d_t)_t$ after differncing

6.2.4 Forecast

In this section, we utilize the previous estimation results of modeling to make the forecast for the WTI oil price data.

Forecast for the Model Fit by the AR(1)+FI(d) Model

To make the forecast in this case, the most important thing is to make the forecast for the volatility ε_t^2. We regard $(\nu_t)_t$ as a Gaussian white noise with zero mean and generate the volatility by

$$\varepsilon_t^2 = (I - B)^{-\hat{d}} \nu_t.$$

In Figure 6.48 ($h = 1$), Figure 6.50 ($h = 2$), Figure 6.52 ($h = 3$), Figure 6.54 ($h = 4$), Figure 6.56 ($h = 5$), Figure 6.58 ($h = 6$) and Figure 6.60 ($h = 7$), we can observe the h-step-ahead forecasts ($h = 1, \cdots, 7$) for the volatility. Then we make the autoregressive regression to obtain the h-step-ahead predictions for Z_t. The results are resented in Figure 6.49 ($h = 1$), Figure 6.51 ($h = 2$), Figure 6.53 ($h = 3$), Figure 6.55 ($h = 4$), Figure 6.57 ($h = 5$), Figure 6.59 ($h = 6$) and Figure 6.61 ($h = 7$). In Table 6.3, we evaluate the forecast behavior by the bias and the RMSE.

Observing the graphs, we find that the forecasts for $(\varepsilon_t^2)_t$ and $(Z_t)_t$ are not far away from the mean. But they do not grasp very well the local volatilities. And the bias and the RMSE are relatively a little too large.

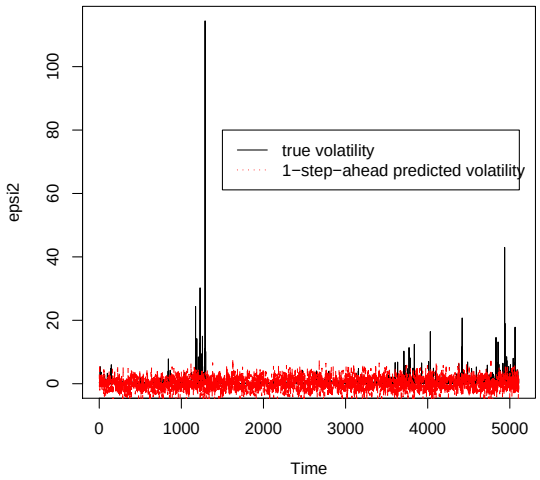


Figure 6.48: WTI: The 1-step-ahead prediction of $(\varepsilon_t^2)_t$ concerning AR(1)+FI(d) model

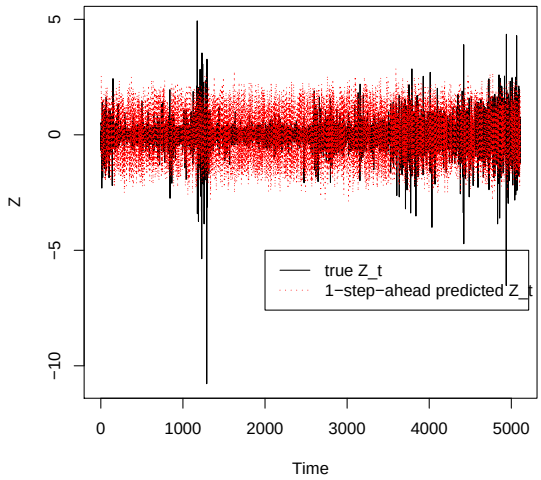


Figure 6.49: WTI: The 1-step-ahead prediction Z_t using AR(1)+FI(d) model

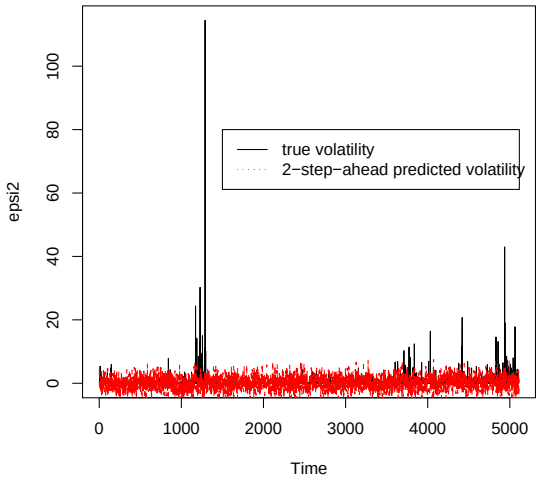


Figure 6.50: WTI: The 2-step-ahead prediction of $(\varepsilon_t^2)_t$ concerning AR(1)+FI(d) model

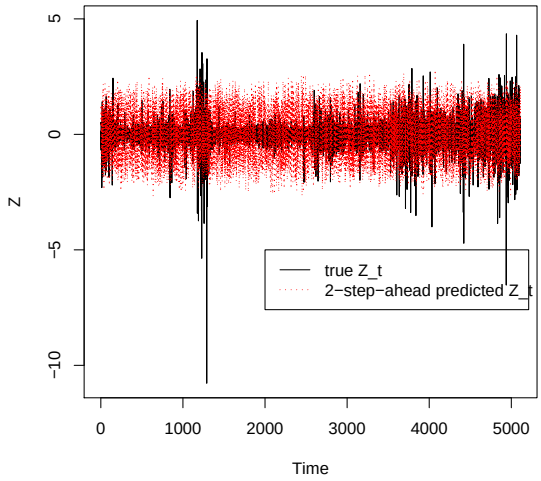


Figure 6.51: WTI: The 2-step-ahead prediction of $(Z_t)_t$ concerning AR(1)+FI(d) model

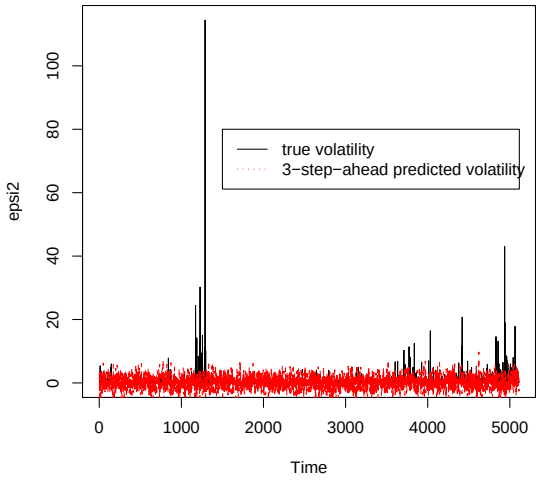


Figure 6.52: WTI: The 3-step-ahead prediction of $(\varepsilon_t^2)_t$ concerning AR(1)+FI(d) model

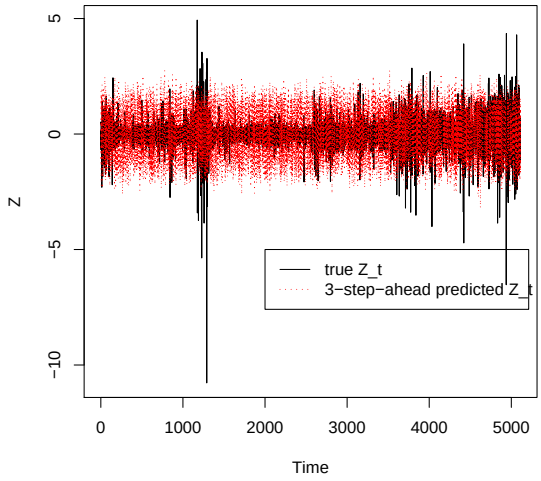


Figure 6.53: WTI: The 3-step-ahead prediction Z_t using AR(1)+FI(d) model

	$h=1$	$h=2$	$h=3$	$h=4$	$h=5$	$h=6$	$h=7$
Bias	6489.46	6503.977	6451.939	6451.207	6377.762	6431.554	6509.753
RMSE	105.2755	74.59163	60.41483	52.38439	46.2088	42.57117	39.73764

Table 6.3: The relative results of the h-step-ahead predictions on Z_t of the WTI oil price data using the AR(1)+FI(d) model.

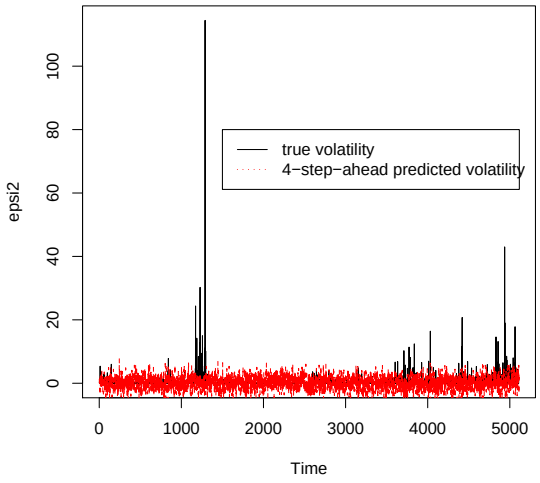


Figure 6.54: WTI: The 4-step-ahead prediction of $(\varepsilon_t^2)_t$ concerning AR(1)+FI(d) model

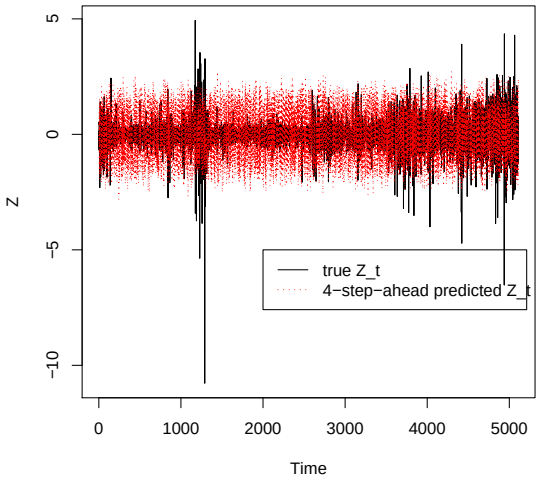


Figure 6.55: WTI: The 4-step-ahead prediction of $(Z_t)_t$ concerning AR(1)+FI(d) model

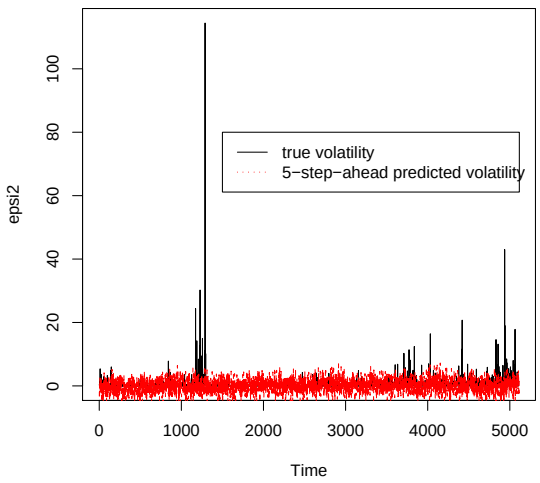


Figure 6.56: WTI: The 5-step-ahead prediction of $(\varepsilon_t^2)_t$ concerning AR(1)+FI(d) model

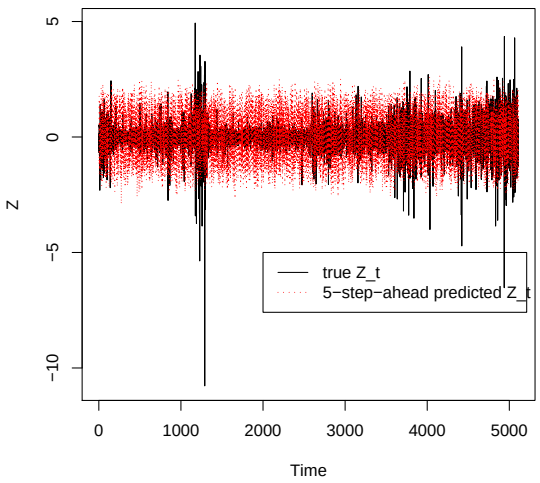


Figure 6.57: WTI: The 5-step-ahead prediction Z_t using AR(1)+FI(d) model

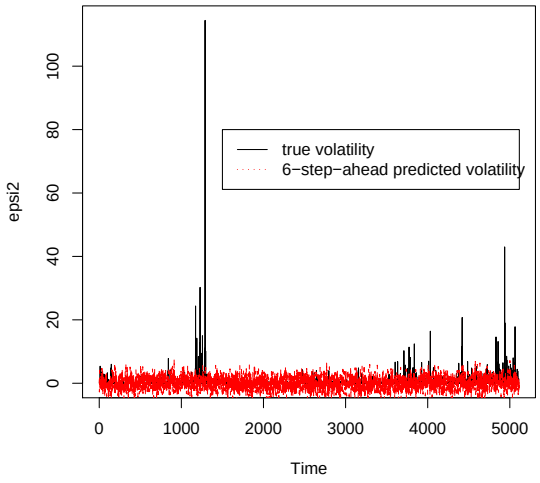


Figure 6.58: WTI: The 6-step-ahead prediction of $(\varepsilon_t^2)_t$ concerning AR(1)+FI(d) model

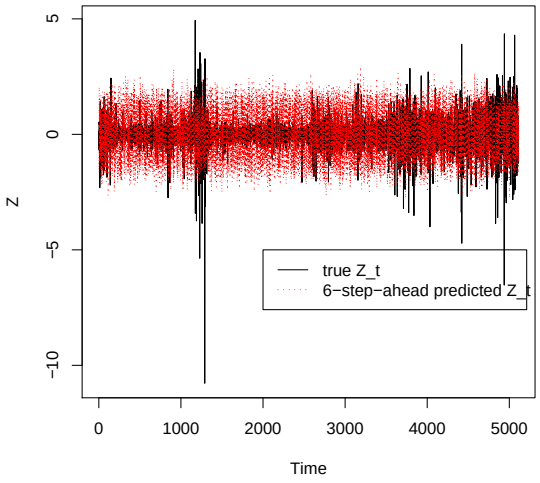


Figure 6.59: WTI: The 6-step-ahead prediction of $(Z_t)_t$ concerning AR(1)+FI(d) model

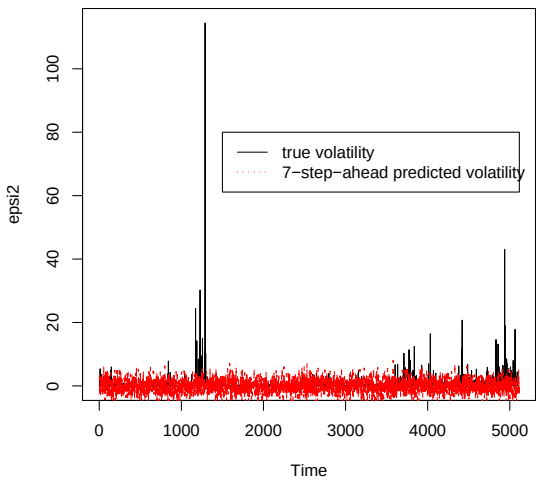


Figure 6.60: WTI: The 7-step-ahead prediction of $(\varepsilon_t^2)_t$ concerning AR(1)+FI(d) model

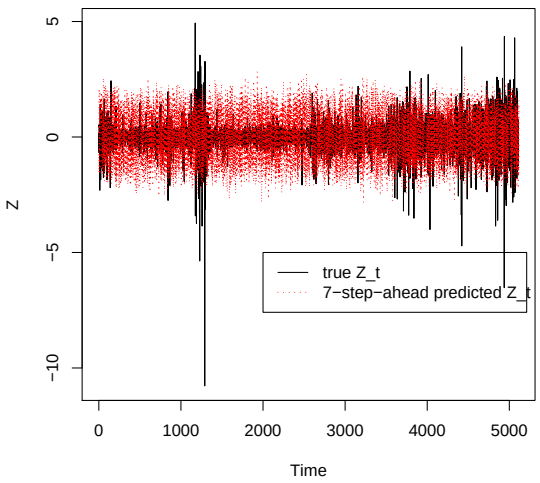


Figure 6.61: WTI: The 7-step-ahead prediction Z_t using AR(1)+FI(d) model

	$h=1$	$h=2$	$h=3$	$h=4$	$h=5$	$h=6$	$h=7$
Bias	6427.122	6396.929	6493.92	6455.806	6425.974	6517.23	6412.771
RMSE	103.7263	73.4362	60.59561	52.3697	46.75349	43.07525	39.19955

Table 6.4: The relative results of the h-step-ahead predictions on Z_t of the WTI oil price data using the AR(2)+FI(d) model.

Forecast for the Model Fit by the AR(2)+FI(d) model

To make the forecast in this case, the most important thing is to make the forecast for the volatility ε_t^2. We regard $(\nu_t)_t$ as a Gaussian white noise with zero mean and generate the volatility by

$$\varepsilon_t^2 = (I - B)^{-\hat{d}}\nu_t.$$

In Figure 6.62 ($h = 1$), Figure 6.64 ($h = 2$), Figure 6.66 ($h = 3$), Figure 6.68 ($h = 4$), Figure 6.70 ($h = 5$), Figure 6.72 ($h = 6$) and Figure 6.74 ($h = 7$), we can observe the h-step-ahead forecasts ($h = 1, \cdots, 7$) for the volatility. Then we make the autoregressive regression to obtain the h-step-ahead predictions for Z_t. The results are resented in Figure 6.63 ($h = 1$), Figure 6.65 ($h = 2$), Figure 6.67 ($h = 3$), Figure 6.69 ($h = 4$), Figure 6.71 ($h = 5$), Figure 6.73 ($h = 6$) and Figure 6.75 ($h = 7$). In Table 6.4, we evaluate the forecast behavior by the bias and the RMSE.

Similarly, observing the graphs, we find that the forecasts for $(\varepsilon_t^2)_t$ and $(Z_t)_t$ are not far away from the mean. But they do not grasp very well the local volatilities. And the bias and the RMSE are relatively a little too large.

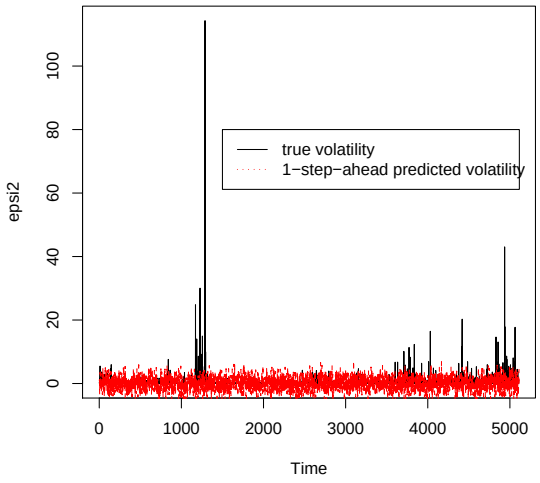


Figure 6.62: WTI: The 1-step-ahead prediction of $(\varepsilon_t^2)_t$ concerning AR(2)+FI(d) model

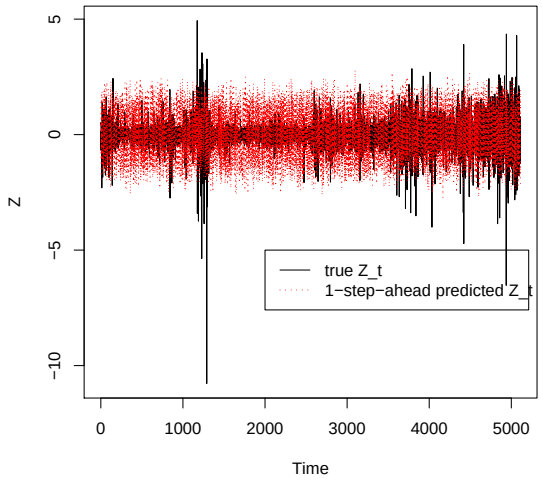


Figure 6.63: WTI: The 1-step-ahead prediction Z_t using AR(2)+FI(d) model

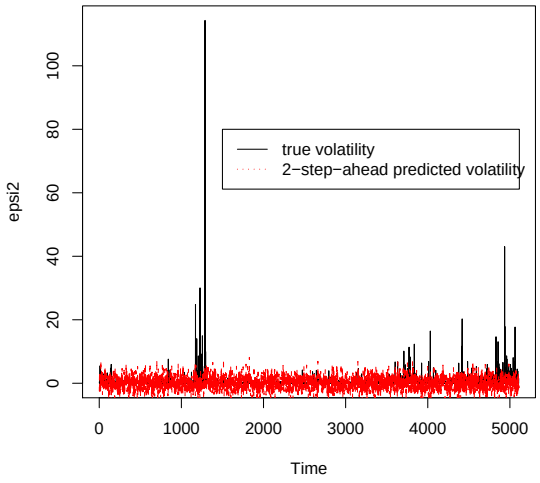


Figure 6.64: WTI: The 2-step-ahead prediction of $(\varepsilon_t^2)_t$ concerning AR(2)+FI(d) model

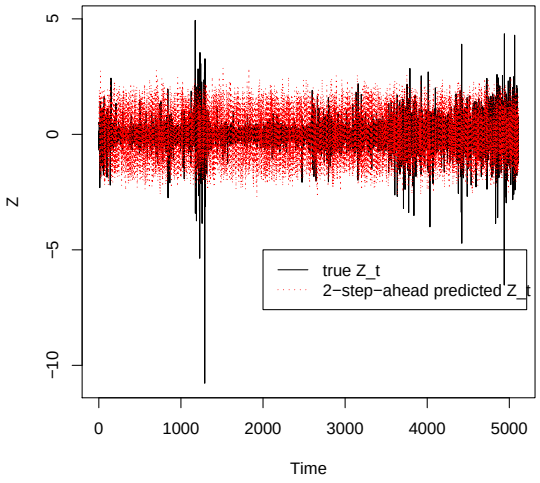


Figure 6.65: WTI: The 2-step-ahead prediction of $(Z_t)_t$ concerning AR(2)+FI(d) model

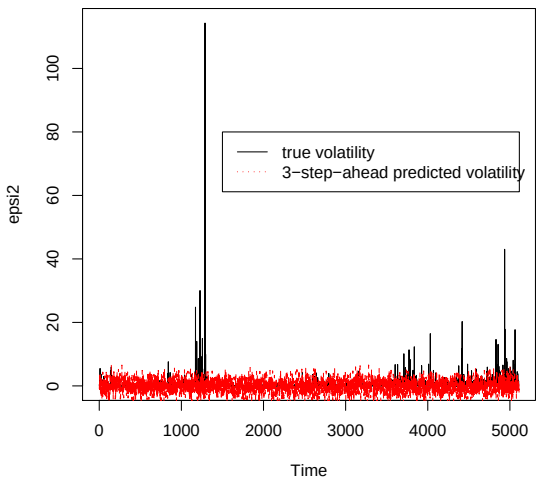


Figure 6.66: WTI: The 3-step-ahead prediction of $(\varepsilon_t^2)_t$ concerning AR(2)+FI(d) model

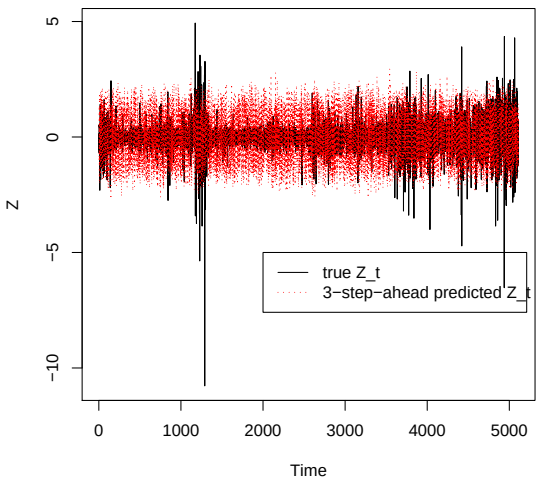


Figure 6.67: WTI: The 3-step-ahead prediction Z_t using AR(2)+FI(d) model

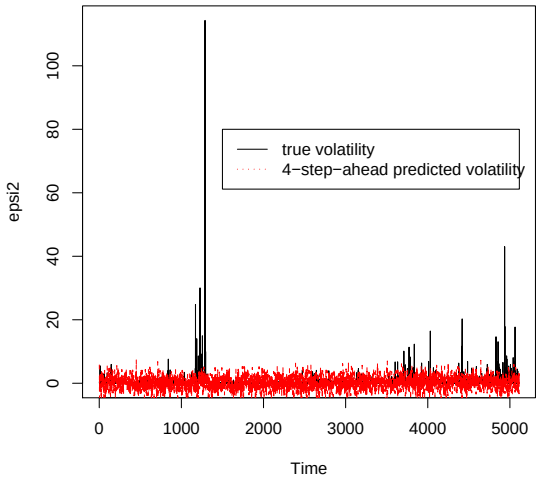


Figure 6.68: WTI: The 4-step-ahead prediction of $(\varepsilon_t^2)_t$ concerning AR(2)+FI(d) model

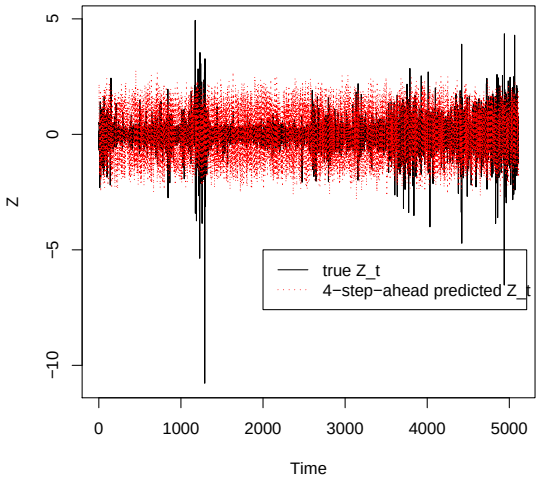


Figure 6.69: WTI: The 4-step-ahead prediction Z_t using AR(2)+FI(d) model

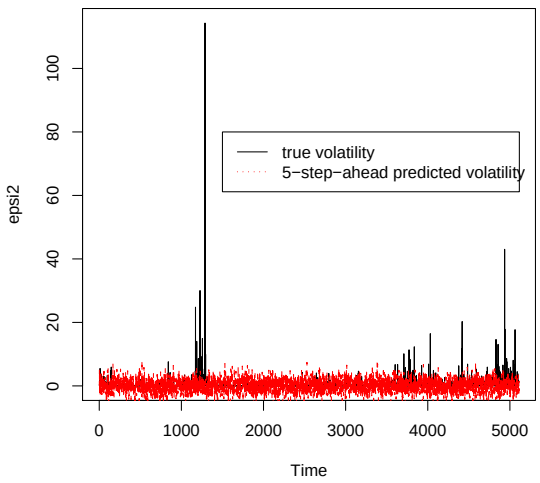


Figure 6.70: WTI: The 5-step-ahead prediction of $(\varepsilon_t^2)_t$ concerning AR(2)+FI(d) model

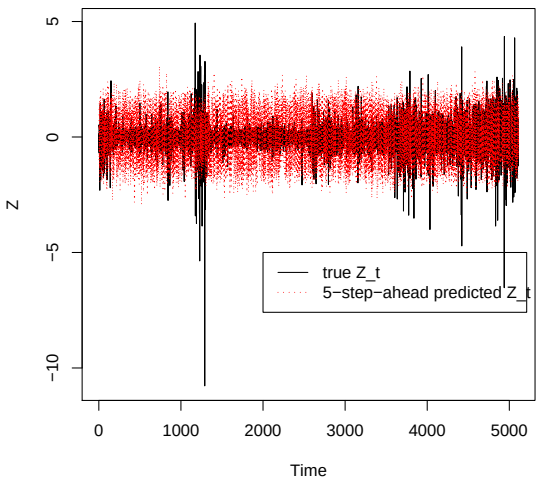


Figure 6.71: WTI: The 5-step-ahead prediction Z_t using AR(2)+FI(d) model

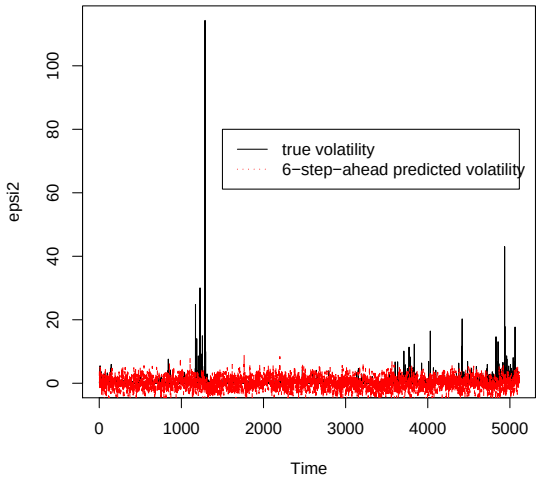


Figure 6.72: WTI: The 6-step-ahead prediction of $(\varepsilon_t^2)_t$ concerning AR(2)+FI(d) model

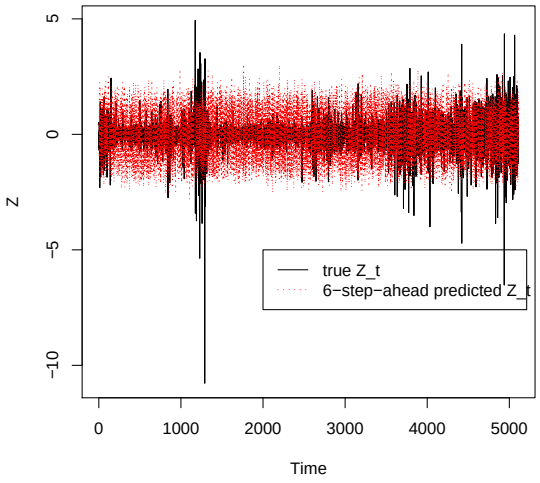


Figure 6.73: WTI: The 6-step-ahead prediction Z_t using AR(2)+FI(d) model

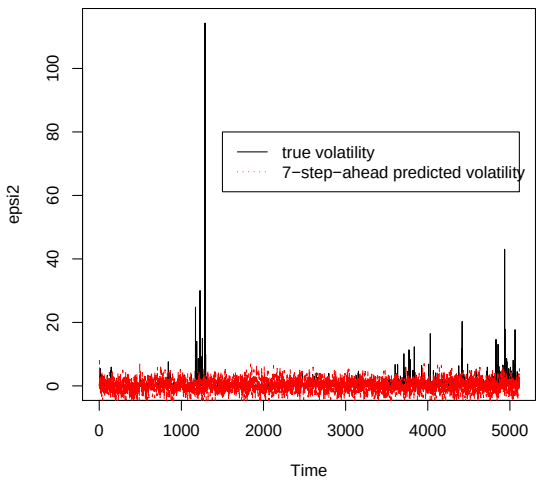


Figure 6.74: WTI: The 7-step-ahead prediction of $(\varepsilon_t^2)_t$ concerning AR(2)+FI(d) model

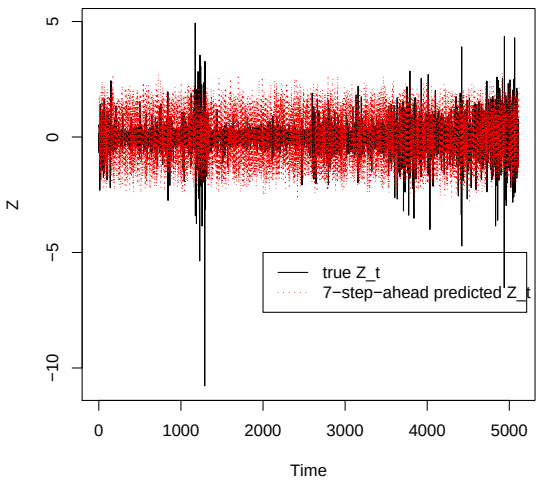


Figure 6.75: WTI: The 7-step-ahead prediction Z_t using AR(2)+FI(d) model

	$h=1$	$h=2$	$h=3$	$h=4$	$h=5$	$h=6$	$h=7$
Bias	6608.34	6983.16	9299.67	6620.821	6360.43	6342.72	8446.96
RMSE	135.25	102.40	110.10	68.47	56.24	53.12	65.91

Table 6.5: The relative results of the h-step-ahead predictions of WTI oil price data using the locally stationary Gegenbauer model (parameter function smoothed by loess method).

Forecast for the Model Fit by the Locally Stationary Gegenbauer Model

Observing the figures of the estimation of the parameter function, we find that without differencing of the original data, the most part of the estimated parameter functions (smoothed by spline and loess method) lie outside the definition interval of the locally stationary Gegenabuer model.

To overcome this awkwardness, we have done some differencing. Thus, the estimation smoothed by loess method lies well in the definition interval of the locally stationary Gegenbauer process, while the estimation smoothed by spline method does not lie totally in the definition interval. In the following, we will make the predictions using the parameter function smoothed by the loess method satisfying the model 6.6.

We first make the h-step-ahead forecasts for $\hat{d}(t)$ ($h = 1, \cdots, 7$) where the parameter function is smoothed by the loess method. We can observe the predictions for $\hat{d}(t)$ in Figure 6.76 ($h = 1$), Figure 6.78 ($h = 2$), Figure 6.80 ($h = 3$), Figure 6.82 ($h = 4$), Figure 6.84 ($h = 5$), Figure 6.86 ($h = 6$) and Figure 6.88 ($h = 7$). Then we utilize these results to forecast the error correction term Z_t, using the strategy described in the previous Chapter. We present in Figure 6.77, Figure 6.79, Figure 6.81, Figure 6.83, Figure 6.85, Figure 6.87 and Figure 6.89 for all the predictions.

From the graphs, we can find that the forecasts for Z_t are to some extent quite good, capturing well the local changes in the trajectory. To have a more clear idea of our forecast, we refer to the two criteria of the forecast. In Table 6.5, the bias and the RMSE of the h-step-ahead predictions are presented. We find that the bias and the RMSE are not so large and thus acceptable, which indicates that the forecast for the model 6.6 is effective.

In fact, if we make the forecast based on the model 6.5, we can imagine the poor result of the forecast due to the fact that the range of the estimate is $[0, 1]$, which is not permitted by the definition of the locally stationary k-factor Gegenbauer process (see Proposition 5.3.1). Thus, if we go on doing the forecast using the estimation based on the locally stationary model without doing any treatments beforehand, the result would not be good. However, if we notice this problem before prediction, we can carry out some differencing such that the parameter function can lie well in the definition interval. That is why we obtained good forecast results based on the model **??**.

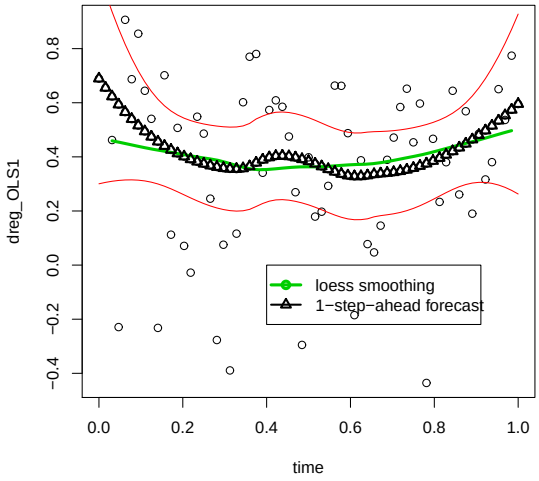


Figure 6.76: 1-step-ahead forecast of the long memory parameter for the WTI data (smoothed by the loess method).

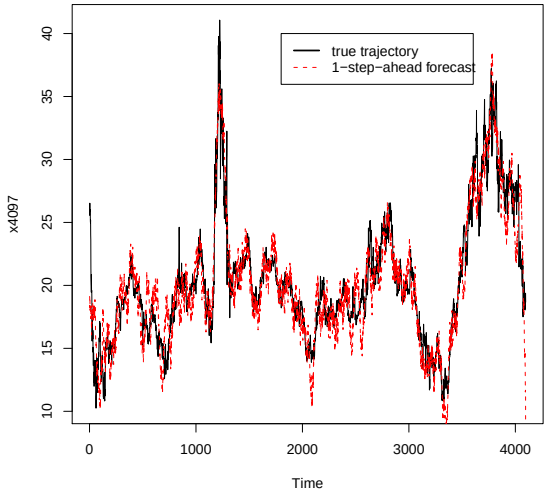


Figure 6.77: 1-step-ahead forecast of the WTI oil price data (smoothed by the loess method).

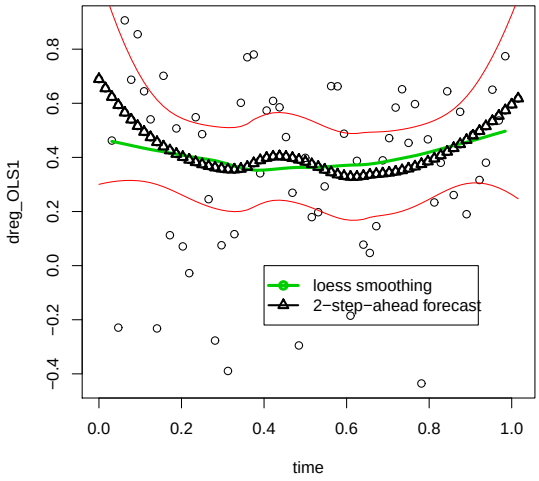


Figure 6.78: 2-step-ahead forecast of the long memory parameter for the WTI data (smoothed by the loess method).

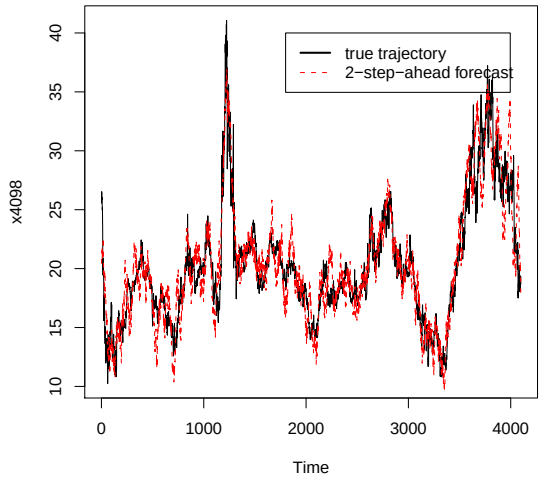


Figure 6.79: 2-step-ahead forecast of the WTI oil price data (smoothed by the loess method).

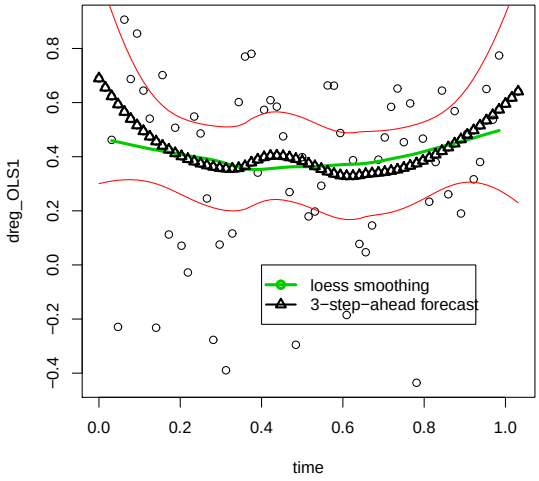


Figure 6.80: 3-step-ahead forecast of the long memory parameter for the WTI data (smoothed by the loess method).

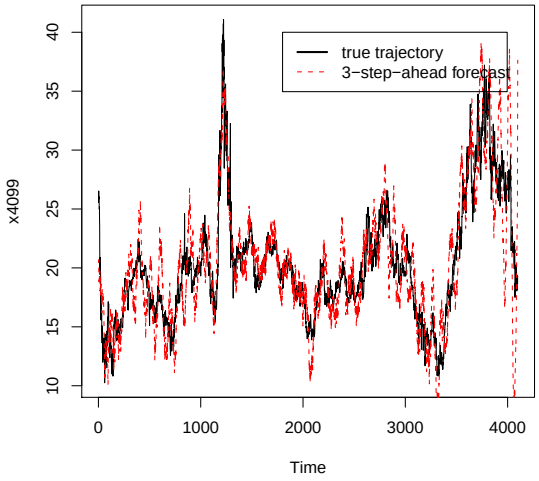


Figure 6.81: 3-step-ahead forecast of the WTI oil price data (smoothed by the loess method).

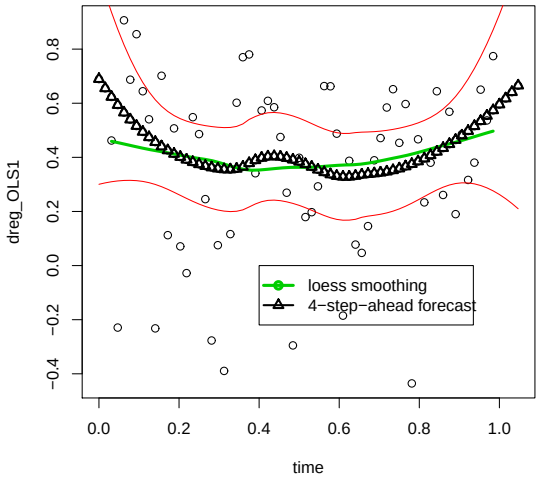


Figure 6.82: 4-step-ahead forecast of the long memory parameter for the WTI data (smoothed by the loess method).

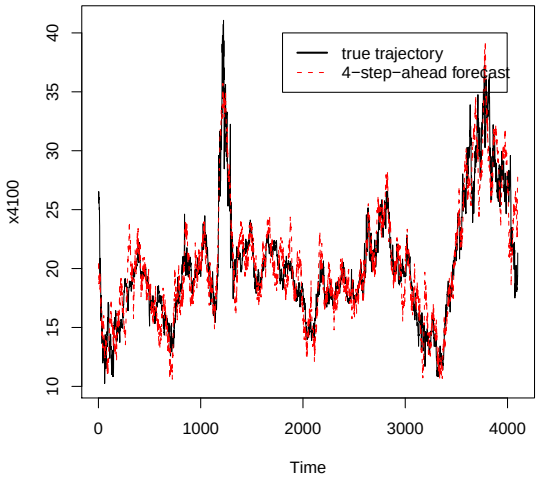


Figure 6.83: 4-step-ahead forecast of the WTI oil price data (smoothed by the loess method).

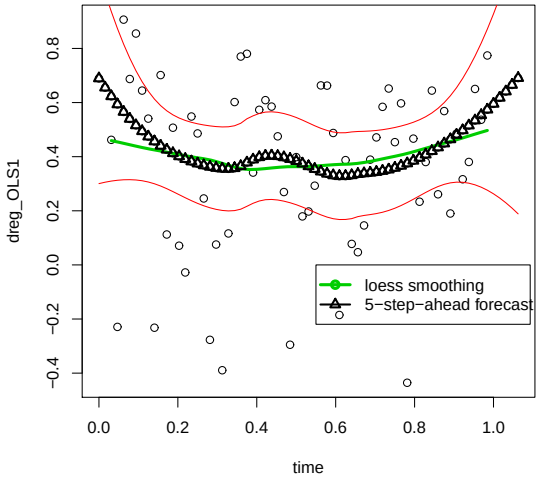


Figure 6.84: 5-step-ahead forecast of the long memory parameter for the WTI data (smoothed by the loess method).

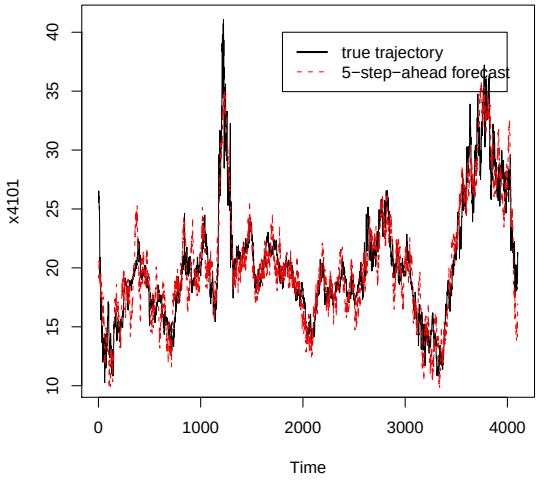


Figure 6.85: 5-step-ahead forecast of the WTI oil price data (smoothed by the loess method).

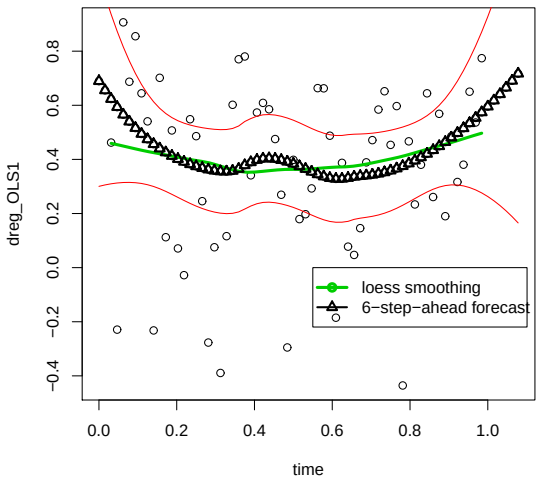


Figure 6.86: 6-step-ahead forecast of the long memory parameter for the WTI data (smoothed by the loess method).

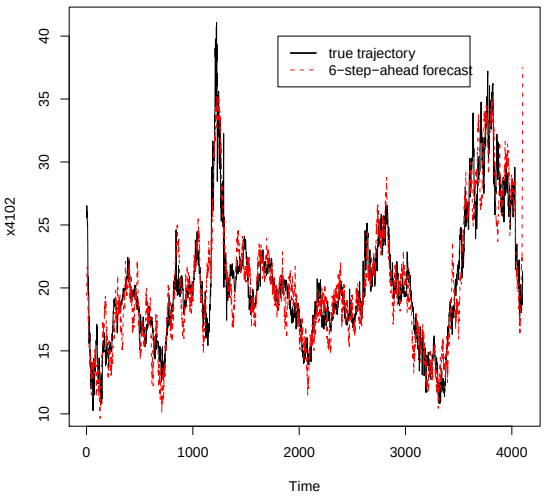


Figure 6.87: 6-step-ahead forecast of the WTI oil price data (smoothed by the loess method).

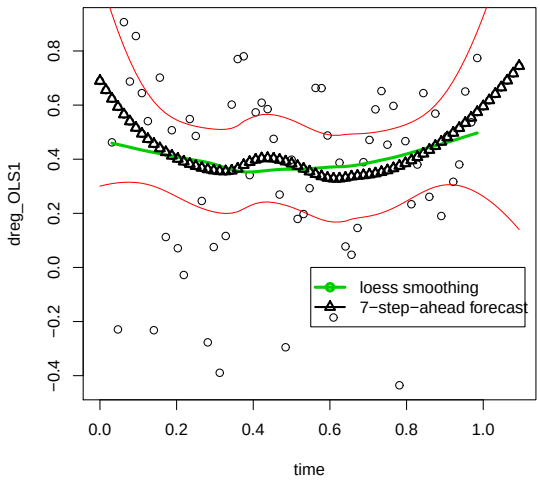


Figure 6.88: 7-step-ahead forecast of the long memory parameter for the WTI data (smoothed by the loess method).

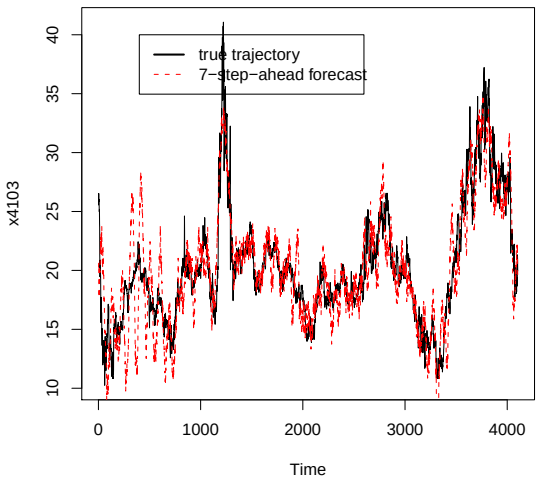


Figure 6.89: 7-step-ahead forecast of the WTI oil price data (smoothed by the loess method).

6.3 Conclusion

In practice, we need to pay attention to the applicability of our model, particularly the range of the definition interval for the parameter function.

If the estimated curve of the long memory parameter function lies well in the interval of the definition of the locally stationary k-factor Gegenbauer process, we can go on carrying out the forecast, and the result will be quite satisfying, for example, the case of NSA225 index data. Otherwise, if the estimated curve is outside the definition interval, we have to do some transformations before modeling.

As we can find in the case of the WTI oil price data, if we apply directly the locally stationary 1-factor Gegenbauer processes, we find that the estimated function is not well located in the definition interval of the model, which will lead to the poor prediction results. However, after suitable differencing, we find that the estimated parameter function smoothed by loess method lies well in the definition interval. Therefore, the graphical result and numerical results are good with no doubt.

Chapter 7

Testing the Fractional Order of Long Memory Processes

Usefulness of fractionally integrated processes has been pointed out over the recent years in order to take various strong persistence effects into modeling. Macroeconomics work on the modeling and forecasting of economic activity time series have made an extensive use of the fractional alternative through long memory processes. For example, Carlin and Dempster (1989) considered monthly unemployment rate of US males; Porter-Hudak (1990) dealt with the US money supply and monetary aggregates and Ray (1993) proposed models for monthly IBM revenue data. Monthly UK inflation rates have been considered by Franses and Ooms (1997), Arteche and Robinson (2000) and Arteche (2003). Other applications also dealt with time series on consumer goods (Darné, Guiraud and Terraza, 2004), public transportation (Ferrara and Guégan, 2000), exchange rates (Ferrara and Guégan, 2001a), spot prices (Ferrara and Guégan, 2001b) or electricity prices (Diongue and Guégan, 2004).

For all those applications, a specific fractionally integrated process has been proposed by the researchers. Generally the choice of the process corresponds to a specific problematic and no comparison has been carried out with different types of long memory processes. For example, some papers used the classical fractionally integrated process, introduced by Granger and Joyeux (1980) and Hosking (1981), while others focused on generalized long memory processes or seasonal long memory processes dedicated to take cyclical or seasonal components with persistence into account.

As we have seen the properties of a time series depend on its order of integration, d, that is on the presence of unit roots. It is important to have techniques available to determine the actual form of non-stationarity and to distinguish between stochastic and deterministic trends if possible. There is a large literature on testing for unit roots theory.

The need to test economic theories which imply random walks has stimulated a large literature involving the unit root distribution (see Dickey and Fuller (1979, 1981), Evans and Savin (1981, 1984), Sargan and Bhargava (1983), Phillips (1987)). One facet of the unit root literature has concerned weakening the assumption of IID errors. In particular

Phillips (1987) showed that the unit root distribution can be used to test for a random walk if the errors satisfy a strong mixing condition. Unfortunately, this condition may not be justified for some economic time series. For example, dependency greater than allowed in Phillips (1987), is permitted by fractionally integrated models which extend the ARIMA(p, d, q) model to real values of d. Furthermore, studies of fractional integration (Granger and Joyeux (1981), Geweke and Porter-Hudak (1983)) have concluded that some economic time series possess fractional unit roots.

7.1 Unit Root Test for Autoregressive Moving Average Processes

Recently, methods for detecting unit roots in autoregressive and autoregressive- moving average time series have been proposed. The presence of a unit root indicates that the time series is not stationary but that differencing will reduce it to stationarity. The tests proposed to date require specification of the number of autoregressive and moving average coefficients in the model.

A good survey may be found in Dickey et al. (1986), among others. Consider the simple AR(1) model:

$$X_t = \phi X_{t-1} + \varepsilon_t, \tag{7.1}$$

where $y_0 = 0$ and the innovation ε_t is a white noise sequence with constant variance. We can regress X_t on X_{t-1} and then use the standard t-statistic to test the null hypothesis: $H_0 : \phi = \phi_0$. The problem arises because we do not know a priori whether the model is stationary. If $|\phi| < 1$, the AR(1) model is stationary, and the ordinary least squares estimator of ϕ, $\hat{\phi}_{OLS}$, equals the maximum likelihood estimator under normality and follows a normal asymptotic distribution. Furthermore, the statistic given by:

$$t_\phi = \frac{\hat{\phi}_{LS} - \phi_0}{s_{\hat{\phi}}},$$

where $s_{\hat{\phi}}$ is the estimated standard deviation of $\hat{\phi}_{LS}$, follows an asymptotic distribution $N(0, 1)$. For small samples, this statistic is distributed approximately as a Student's t with $(T - 1)$ degrees of freedom. Nevertheless, when $\phi = 1$, this result does not hold. It can be shown that the OLS estimator of ϕ is biased downwards and that the t-statistic under the unit-root null hypothesis, does not have a Student's t distribution even in the limit as the sample size becomes infinite. The AR(1) model (7.1) can be written as follows by subtract X_{t-1} at both sides of the equation:

$$X_t - X_{t-1} = (\phi - 1)X_{t-1} + \varepsilon_t \tag{7.2}$$

or

$$\Delta X_t = \rho X_{t-1} + \varepsilon_t, \tag{7.3}$$

where $\rho = \phi - 1$. The relevant unit root hypothesis is $\rho = 0$ and the alternative is one side:

$$H_a : \rho < 0,$$

since $\rho > 0$ corresponds to explosive time series models. Dickey (1976) tabulated the percentiles of this statistic under the unit root null hypothesis. The H_0 of a unit root is rejected when the value of the statistic is lower than the critical value. This statistic, denoted by τ, is called the Dickey-Fuller statistic and their critical values are published in Fuller (1976).

Up to now it has been shown how to test the null hypothesis of a random walk (one unit root) against the alternative of a zero mean, stationary . For economic time series, it could be of interest to consider alternative hypothesis including stationarity around a constant and/or a linear trend. This could be achieved by introducing these terms in model (7.3):

$$\Delta X_t = \alpha + \rho X_{t-1} + \varepsilon_t, \tag{7.4}$$

$$\Delta X_t = \alpha + \beta t + \rho X_{t-1} + \varepsilon_t. \tag{7.5}$$

The unit root null hypothesis is simply $H_0 : \rho = 0$ in both model (7.4)-(7.5). Dick-Fuller tabulated the critical values for the corresponding statistics, denoted by τ_μ and τ_τ respectively. It should be noted that model (7.5) under the null hypothesis becomes a random walk plus drift model, which is a hypothesis that frequently arises in economic applications.

The augmented Dickey-Fuller (ADF) test tests the null hypothesis that a time series X_t is FI(1) against the alternative that it is FI(0), assuming that the dynamics in the data have an ARMA structure. Hassler and Wolters (1994) found that the Augmented Dickey-Fuller test (hereafter ADF) against fractional alternatives lost considerable power when augmented terms were added. In contrast, Krämer and Dittmann (1998) showed that this test is consistent if the order of the autoregression does not tend to infinity too fast.

7.2 Unit Root Test for Fractional Integrated Processes

The class of fractionally integrated processes, denoted as FI(d), where the order of integration d is extended to be any real number, has proved very useful in capturing the persistence properties of many long-memory processes: Baillie (1996), Beran (1994), and Granger and Joyeux (1980). In general, unit root tests are consistent when the alternative is a FI(d) process but their power turns out to be quite low (see Diebold and Rudebusch (1991), Schmidt and Lee (1996)). In particular, this lack of power has motivated the development of new testing approaches that take this type of alternative explicitly into consideration. There is a growing literature on this subject that can be basically classified into two strands. First, there are Wald-type tests that, by working under the alternative hypothesis, provide point estimates of the memory parameter and build confidence intervals around it. Secondly, there are Lagrange Multiplier (LM) tests where statistics are evaluated under the corresponding null hypothesis. Within the first group, there are a very large number of rather heterogeneous contributions: parametric and semiparametric methods of estimating d both in the frequency and in the time domain (see, inter

alia, Geweke and Porter-Hudak (1983), Fox and Taqqu (1986), Sowell (1992), Robinson (1992)). However, most of them lack power when used for testing purposes. On one hand, the semiparametric techniques tend to yield large confidence intervals that include too often the null hypothesis. On the other hand, although in general the parametric methods present narrower confidence intervals, the precision with which the parameters are estimated hinges on the correct specification of the model (see Hauser, Potscher and Reschenhofer (1999)). Within the second group, Robinson (1994) and Tanaka (1999) have proposed useful LM tests in the frequency and the time domain, respectively. A distinctive feature of both approaches is that, in contrast to the classical unit root tests where asymptotic distributions are nonstandard and require case-by-case numerical tabulation, they do have standard asymptotic distributions. In this respect, Robinson (1994) has attributed this different limit behavior to the use of an explicit autoregressive (AR) alternative in the classical unit root-testing approach. Nonetheless, despite the advantage of having a standard limit distribution, a possible shortcoming of the LM approach is that, by working under the null hypothesis, it does not yield any direct information about the correct long memory parameter, d, when the null is rejected.

In order to overcome that drawback, Dolado et al. (2002) proposed a simple Wald-type test in the time domain that has acceptable power properties and, as a by-product of its implementation, provides information about the values of d under the alternative hypothesis. It turns out to be a generalization of the well-known Dickey-Fuller (D-F) test, originally developed for the case of FI(1) versus FI(0), to the more general case of FI(d_0) versus FI(d_1) with $d_1 < d_0$ and, thus, they referred to it as the Fractional Dickey-Fuller (FD-F) test. When $d_0 = 1$, the proposed test statistics are based on the OLS estimator, or its t-ratio, of the coefficient on $\Delta^{d_1} X_{t-1}$ in a regression of ΔX_t on $\Delta^{d_1} X_{t-1}$ and, possibly, some lags of ΔX_t. When d_1 is not taken to be known a priori, a pre-estimation of d, is needed to implement the test. We show that the choice of any $T^{1/2}$-consistent estimator of $d_1 \in [0, 1)$ suffices to make the test feasible, while achieving asymptotic normality.

Another well-known unit root test was proposed by Robinson (1994). He investigated a general model in order to test whether the data stemmed from a stationary or a non-stationary process, under uncorrelated and weakly-correlated innovations $(\varepsilon_t)_t$. The process described by the following equation nests all the specific long memory processes generally used in applications:

$$F(B)X_t = (I - B)^{d_0+\theta_0} \prod_{i=1}^{k-1} (I - 2\nu_i B + B^2)^{d_i+\theta_i} (I + B)^{d_k+\theta_k} X_t = \varepsilon_t, \qquad (7.6)$$

where B is the backshift operator. For $i = 1, \cdots, k-1$, $\nu_i = cos\lambda_i$, λ_i being any frequency between 0 and π. For $i = 0, 1, \cdots k$, θ_i belongs to $[-1, 1]$ and d_i is such that: $|d_i| < 1/2$, implying thus that the spectral density is unbounded in λ_i. Moreover, $(\varepsilon_t)_t$ is an innovation process to be specified.

To adjust these models on real data sets, it is fundamental to detect long memory behavior through statistical tests. Here, we investigate Robinson (1994) test. The properties

of this test have been proved in an asymptotic setting. Working with macroeconomics data sets significates that we deal with data sets with rather small sample sizes, generally lower than 1 000 points. Thus it is crucial for practitioners to know the accuracy of this test for a finite sample size. We assess the rate of convergence for Robinson's test using Monte Carlo simulations.

From a practical point of view, before implementing a fractional process on real data, it is warmly recommended to carry out a statistical test to show evidence of persistence in the data. In this respect, the test of Robinson (1994) has been proved to be very useful for testing stationarity of many SCLM processes (see Gil-Alana, 2001, 2006). This test also permits to test the integration order at various frequencies and does not require the estimation of long memory parameters since the series have a short memory behavior under the null hypothesis. Now, this test can also be used to test the degree of persistence of the memory parameter using the null of Robinson test. However, one of the major drawback in empirical macroeconomics is the rather small amount of data available to the practitioners. For example in the industrialized countries, the broadest measure of economic activity released by the quarterly national accounts of the statistical institutes, namely GDP, is generally available only since 1970. Thus, in most of the cases, less than 160 data points are available to carry out the analysis on a quarterly basis. In this respect, it appears crucial to study the finite sample behavior of the statistical procedures carried out at each step of the statistical analysis. In this paper, we propose a simulation experiment to determine the possible application of the Robinson (1994) test for finite samples. Indeed, in his paper the results were proved in an asymptotic setting and we need to know how the test works empirically. We study the convergence of the test according to the fractional process used to generate simulated data. We investigate this rate of convergence whatever the innovation process is uncorrelated or weakly correlated.

We briefly describe Robinson test (1994) which is a Lagrange Multiplier test for testing unit roots and other fractionally hypotheses when the roots are located at any frequency on the interval $[0, \pi]$. The test is derived via the score principle and its asymptotic critical values follow the Chi-squared distribution. Let $(Y_t)_t$ be a stochastic process such that:

$$Y_t = \beta' Z_t + X_t, \tag{7.7}$$

where $(Z_t)_t$ is a $k \times 1$ observable vector, β an unknown $k \times 1$ vector and $(X_t)_t$ a process which follows Equation (7.6). In the rest of the paper, we assume that $\beta = 0$ and $(\varepsilon_t)_t$ is either a strong white noise or a GARCH(1,1) noise.

Robinson (1994) worked with the general model (7.6) for a fixed d and tested the assumption

$$H_0 : \theta = (\theta_0, \cdots, \theta_k)' = 0,$$

against the alternative:

$$H_a : \theta \neq 0.$$

The test statistic is defined by:

$$\tilde{R} = \frac{T}{\tilde{\sigma}^4} \frac{\tilde{a}^2}{\tilde{A}}, \tag{7.8}$$

where T is the length of the raw time series and

$$\tilde{\sigma}^2 = \frac{2\pi}{T} \sum_{j}^{*} I_{\tilde{\varepsilon}}(\lambda_j).$$

$I_{\tilde{\varepsilon}}(\lambda_j)$ is the periodogram of $\tilde{\varepsilon}_t$ with $\tilde{\varepsilon}_t = F(B)Y_t$, $F(B)$ being given in Equation (7.6). Moreover, we get:

$$\tilde{A} = \frac{2}{T} \sum_{j}^{*} \psi(\lambda_j) \cdot \psi(\lambda_j)',$$

and

$$\tilde{a}^2 = \frac{-2\pi}{T} \sum_{j}^{*} \psi(\lambda_j) I(\lambda_j),$$

where $\sum_{j}^{*}$ is the sum over $\lambda_j = \frac{2\pi j}{T} \in M = \{\lambda : -\pi < \lambda < \pi, \lambda \notin (\rho_l - \eta, \rho_l + \eta)\}$ such that ρ_l are the distinct poles of $\psi(\lambda)$ on $(-\pi, \pi]$, η is a given positive constant. Finally, we get:

$$\psi(\lambda_j) = (\psi_l(\lambda_j)),$$

with

$$\psi_l(\lambda_j) = \delta_{0l} \log |2sin\frac{1}{2}\lambda_j| + \delta_{kl} \log(2cos\frac{1}{2}\lambda_j) + \sum_{i=1}^{k} \delta_{il} \log(|2(cos(\lambda_j) - \cos\omega_i)|,$$

for $l = 0, 1, \cdots, k$, where $\delta_{il} = 1$ if $i = l$ and 0 otherwise.

Under stationary conditions, Robinson (1994) established that:

$$\tilde{R} \rightarrow^d \chi^2_{k+1},$$

where $k + 1 = dim(\theta)$. If χ^2_{k+1} represents the χ^2 distribution with $k + 1$ degrees of freedom then $\chi^2_{k+1,\alpha}$ represents a quantile for a given level α. As soon as $\tilde{R} > \chi^2_{k+1,\alpha}$, we reject H_0, with a risk α.

Under the null, the test chooses the best long memory parameter which corresponds to the greatest p-value of the Chi-square test. We accept the null hypotheses if the p-value is greater than the significant level and we reject it if the p-value is smaller than or equal to the significance level. The test appears as a method to test the long memory parameters. We can perform the properties of this test to estimate parameters using a Monte Carlo simulation which provides the mean, bias and RMSE for a suitable number of replications.

Monte Carlo Experiment

In this section we carry out the Monte Carlo experiments for several models derived from Equation (7.6) using different sample sizes with replications.

Under (H_0), we simulate different models using first a strong Gaussian white noise $(\varepsilon_t)_t$ with zero mean and unit variance and second a GARCH (1,1) noise. In that latter case, $\varepsilon_t = \sqrt{h_t}\xi_t$, $h_t = a_0 + a_1\varepsilon_{t-1}^2 + b_1 h_{t-1}$, with $(\xi_t)_t$ being a sequence of i.i.d. Gaussian random variables with zero mean and unit variance, $a_0 = 1$, $a_1 = 0.15$ and $b_1 = 0.8$.

We consider nine various models: one model (4.2) , two models (4.3), two models (4.5) and four models (4.6). For the models (4.3), we use $s = 4$ and $s = 12$, then we mix the possible explosion at frequency zero with the explosions with fixed seasonalities assuming $s = 4$ and $s = 12$. For the Gegenbauer models, we detail the results with respect to the location frequency and the number of explosions inside the spectral density. When we have only one factor in the model, the true value of the long memory parameter is $d = 0.3$ (except for the model $(1 + B)^d = \varepsilon_t$, for which we use $d = 0.2$); in presence of two factors, we use $d_1 = 0.3$ and $d_2 = 0.4$; in presence of three factors, we use $d_1 = 0.2$, $d_2 = 0.3$ and $d_3 = 0.4$.

We consider several sample sizes T from 100 to 3000. We do not give the results up to 3000 because we intend to apply the method to macroeconomic data sets whose sizes are generally smaller. In all cases, we use two sizes of replication, $TT = 100, 1000, 5000$. We only present the results for $TT = 100$, because the results are quite similar with $TT = 1000$ and $TT = 5000$. The results are available upon request.

We carry out the code on the computer Mac OS X 10.5.1 Léopard, written in language R. The random numbers are generated by the command "rnorm()" as the pseudo random numbers. In the tables the notation $\hat{d}$ represents the mean of the TT realisations $(\hat{d}_1, \ldots, \hat{d}_{TT})$ possessing the greatest p-value of the test. In the tables, n represents the percentage of times that we get the true value for all the long memory parameters involved into the models.

We find that for the models with only one term, like models (4.2) and (4.6) with $k = 1$, the test performs correctly for sample size greater than 900. However, for the models (4.3), although there is only one parameter to test, the test does not present good performances. The performances become correct for sample sizes equal to 2000 and 3000. The same results are observed when we simulate models with several factors, like the models (4.5) and (4.6) with $k \geq 2$. The more the explosions are inside the spectral density, the worse is the test's performance. We never got convergence for the test applying at the model (4.7) as soon as $k > 3$. From a general point of view, as expected, the performance of the test increases with the sample size.

The main results are the following:

1. First, we assume that the noise $(\varepsilon_t)_t$ is a strong white noise in all the models:
   - For the model (4.2), $\hat{d} = 0.3$ when T reaches 3000.
   - For the models (4.3), $\hat{d} = 0.3$, for $s = 4$ when T reaches 3000, and the test does not converge when we use $s = 12$.

- For the models (4.5), $\hat{d} = 0.3$, for $s = 4$ and $s = 12$, when T reaches 3000.
- For the 1-factor model (4.6),
  (a) If $\nu = -1$, $\hat{d} = 0.201$ when T reaches 700.
  (b) If $\nu = cos(\pi/3)$, $\hat{d} = 0.3$ when T reaches 2000.
- For a 2-factors model (4.6), $\hat{d}_1 = 0.3$ and $\hat{d}_2 = 0.4$ when T reaches 3000.
- For a 3-factors model (4.6), $\hat{d}_1 = 0.2$, $\hat{d}_2 = 0.3$ and $\hat{d}_3 = 0.4$ when T reaches 2000.

2. Second, we assume that the noise $(\varepsilon_t)_t$ is a GARCH(1,1) noise for all the models:

- For the model (4.2), $\hat{d} = 0.299$ when T reaches 3000.
- For the model (4.3), $\hat{d} = 0.3$, for $s = 4$ when T reaches 3000, and does not converge when we use $s = 12$.
- For the models (4.5), the test does not converge.
- For the 1-factor model (4.6),
  (a) If $\nu = -1$, $\hat{d} = 0.206$ when T reaches 1000.
  (b) If $\nu = cos(\pi/3)$, $\hat{d} = 0.3$ when T reaches 2000.
- For a 2-factors model (4.6), $\hat{d}_1 = 0.3$ and $\hat{d}_2 = 0.4$ when T reaches 3000.
- For a 3-factors model (4.6), $\hat{d}_1 = 0.2$, $\hat{d}_2 = 0.3$ and $\hat{d}_3 = 0.4$ when T reaches 3000.

In the presence of an infinite cycle, when we simulate the models (4.4) and (4.5), the comparison of the performances of the test shows that the convergence is slower. Sometimes the test does not converge at all. We observe also that the test does not converge when we use the model (4.3) with $s = 12$. In any case, the test convergence is very slow for all the models we use. As soon as we have more than one explosion, we need to use almost 1000 data to be sure to attain in mean the correct estimated value. When we have more than one explosion inside the spectral density, it appears difficult to use the test for samples whose size is smaller than 3000. The results are quite similar whatever the noise we use for simulations: a strong white noise or a GARCH noise.

Conclusion

In this part, we evaluate the performances of the Robinson (1994) test for several simulated SCLM models. We show that the sample size is crucial for the accuracy of the test. It appears that the use of this test is mainly recommended when we observe only one explosion in the spectral density if we have at least 500 points. If there exists more than one explosion inside the spectral density, this test does not provide accurate information if the sample size of the data set is less than 3000. This latter result raises concern as regards the applications of Robinson test to seasonal macroeconomics data.

T	100	200	300	400	500	600	700	800	900	1000	2000	3000
n	37	52	67	79	80	85	88	90	94	98	100	100
$\hat{d}$	0.248	0.26	0.283	0.291	0.288	0.289	0.29	0.296	0.302	0.3	0.3	0.3

Table 7.1: Test for model $(1-B)^{0.3}X_t = \varepsilon_t$ where ε_t is a strong white noise.

T	100	200	300	400	500	600	700	800	900	1000	2000	3000
n	7	23	43	50	58	63	72	77	85	92	99	100
$\hat{d}$	0.092	0.187	0.235	0.25	0.258	0.264	0.276	0.277	0.285	0.292	0.299	0.3

Table 7.2: Test for model $(1-B^4)^{0.3}X_t = \varepsilon_t$ where ε_t is a strong white noise.

T	100	200	300	400	500	600	700	800	900	1000	2000	3000
n	0	0	0	0	3	3	4	10	18	26	78	98
$\hat{d}$	0.0	0.005	0.05	0.088	0.148	0.18	0.191	0.204	0.217	0.2223	0.278	0.298

Table 7.3: Test for model $(1-B^{12})^{0.3}X_t = \varepsilon_t$ where ε_t is a strong white noise.

T	100	200	300	400	500	600	700	800	900	1000	2000	3000
n	3	14	17	26	28	38	50	52	51	63	87	92
$\hat{d}_1$	0.279	0.314	0.296	0.307	0.305	0.297	0.299	0.307	0.299	0.3	0.3	0.302
$\hat{d}_2$	0.154	0.287	0.332	0.338	0.346	0.36	0.372	0.376	0.377	0.386	0.396	0.399

Table 7.4: Test for model $(1-B)^{0.3}(1-B^4)^{0.4}X_t = \varepsilon_t$ with $d_1 = 0.3$ where ε_t is a strong white noise.

T	100	200	300	400	500	600	700	800	900	1000	2000	3000
n	0	0	3	1	4	7	10	16	20	27	71	92
$\hat{d}_1$	0.299	0.283	0.304	0.303	0.304	0.309	0.298	0.306	0.302	0.307	0.3	0.301
$\hat{d}_2$	0.001	0.019	0.137	0.175	0.245	0.272	0.284	0.303	0.322	0.329	0.373	0.394

Table 7.5: Test for model $(1-B)^{0.3}(1-B^{12})^{0.4}X_t = \varepsilon_t$ where ε_t is a strong white noise.

T	100	200	300	400	500	600	700	800	900	1000	2000	3000
n	34	59	65	78	78	91	89	97	94	98	100	100
$\hat{d}$	0.246	0.286	0.285	0.292	0.286	0.293	0.297	0.299	0.3	0.3	0.3	0.3

Table 7.6: Test for model $(1-2\nu B+B^2)^{0.15}X_t = (1+B)^{0.3} = \varepsilon_t$ where ε_t is a strong white noise and $\nu = -1$.

T	100	200	300	400	500	600	700	800	900	1000	2000	3000
n	32	55	65	74	75	90	87	90	97	94	99	100
$\hat{d}$	0.159	0.17	0.179	0.19	0.189	0.196	0.201	0.194	0.201	0.196	0.201	0.2

Table 7.7: Test for model $(1-2\nu B+B^2)^{0.1}X_t = (1+B)^{0.2} = \varepsilon_t$ where ε_t is a strong white noise and $\nu = -1$.

T	100	200	300	400	500	600	700	800	900	1000	2000	3000
n	18	42	60	69	78	91	94	94	98	95	100	100
$\hat{d}$	0.171	0.235	0.266	0.272	0.278	0.293	0.294	0.296	0.298	0.297	0.3	0.3

Table 7.8: Test for model $(1-2\nu B+B^2)^{0.3}X_t = \varepsilon_t$ where ε_t is a strong white noise, $\nu = \cos\dfrac{\pi}{3}$.

T	100	200	300	400	500	600	700	800	900	1000	2000	3000
n	0	7	23	29	32	47	52	56	69	67	94	100
$\hat{d}_1$	0.068	0.152	0.192	0.225	0.231	0.255	0.254	0.267	0.277	0.28	0.295	0.3
$\hat{d}_2$	0.223	0.29	0.322	0.343	0.351	0.364	0.372	0.378	0.388	0.389	0.398	0.4

Table 7.9: Test for model $(1-2\nu_1 B+B^2)^{0.3}(1-2\nu_2 B+B^2)^{0.4}X_t = \varepsilon_t$ where ε_t is a strong white noise, $\nu_1 = \cos\dfrac{\pi}{3}, \nu_2 = \cos\dfrac{5\pi}{6}$.

T	100	200	300	400	500	600	700	800	900	1000	2000	3000
n	0	4	8	21	33	50	58	63	77	73	100	100
$\hat{d}_1$	0.023	0.085	0.129	0.142	0.156	0.173	0.176	0.18	0.192	0.187	0.2	0.2
$\hat{d}_2$	0.16	0.229	0.245	0.25	0.271	0.278	0.284	0.285	0.295	0.286	0.3	0.3
$\hat{d}_3$	0.133	0.247	0.306	0.33	0.353	0.369	0.369	0.372	0.388	0.385	0.4	0.4

Table 7.10: Test for model $(1 - 2\nu_1 B + B^2)^{0.2}(1 - 2\nu_2 B + B^2)^{0.3}(1 - 2\nu_3 B + B^2)^{0.4} X_t = \varepsilon_t$ where ε_t is a strong white noise, $\nu_1 = \cos\frac{\pi}{6}, \nu_2 = \cos\frac{\pi}{2}, \nu_3 = \cos\frac{2\pi}{3}$.

T	100	200	300	400	500	600	700	800	900	1000	2000	3000
n	34	50	56	63	72	76	80	81	83	84	99	99
$\hat{d}$	0.249	0.278	0.279	0.281	0.292	0.291	0.296	0.291	0.301	0.298	0.299	0.299

Table 7.11: Test for model $(1 - B)^{0.3} X_t = \varepsilon_t$ where ε_t is GARCH(1,1).

T	100	200	300	400	500	600	700	800	900	1000	2000	3000
n	6	24	38	46	43	73	64	72	81	72	91	98
$\hat{d}$	0.112	0.193	0.235	0.251	0.244	0.274	0.27	0.274	0.281	0.272	0.291	0.3

Table 7.12: Test for model $(1 - B^4)^{0.3} X_t = \varepsilon_t$ where ε_t is GARCH(1,1).

T	100	200	300	400	500	600	700	800	900	1000	2000	3000
n	0	0	1	0	7	9	12	13	20	29	75	94
$\hat{d}$	0.0	0.005	0.06	0.11	0.156	0.182	0.192	0.204	0.218	0.229	0.275	0.294

Table 7.13: Test for model $(1-B^{12})^{0.3}X_t = \varepsilon_t$ where ε_t is GARCH(1,1).

T	100	200	300	400	500	600	700	800	900	1000	2000	3000
n	0	4	14	20	22	32	36	49	50	54	83	86
$\hat{d}_1$	0.283	0.312	0.297	0.292	0.312	0.289	0.304	0.3	0.3	0.298	0.306	0.299
$\hat{d}_2$	0.152	0.257	0.319	0.335	0.34	0.369	0.372	0.378	0.381	0.385	0.393	0.395

Table 7.14: Test for model $(1-B)^{0.3}(1-B^4)^{0.4}X_t = \varepsilon_t$ where ε_t is GARCH(1,1).

T	100	200	300	400	500	600	700	800	900	1000	2000	3000
n	0	0	0	2	5	9	8	13	22	20	63	81
$\hat{d}_1$	0.294	0.294	0.28	0.302	0.298	0.302	0.299	0.305	0.303	0.305	0.304	0.304
$\hat{d}_2$	0	0.02	0.117	0.177	0.249	0.278	0.285	0.298	0.314	0.323	0.37	0.386

Table 7.15: Test for model $(1-B)^{0.3}(1-B^{12})^{0.4}X_t = \varepsilon_t$ where ε_t is GARCH(1,1).

T	100	200	300	400	500	600	700	800	900	1000	2000	3000
n	31	49	59	66	71	76	84	85	86	86	94	100
$\hat{d}$	0.232	0.273	0.294	0.284	0.291	0.298	0.304	0.299	0.299	0.3	0.3	0.3

Table 7.16: Test for model $(1-2\nu B+B^2)^{0.15} = (1+B)^{0.3}X_t = \varepsilon_t$ where ε_t is GARCH(1,1) and $\nu = -1$.

T	100	200	300	400	500	600	700	800	900	1000	2000	3000
n	36	54	65	61	67	73	82	87	88	86	94	99
$\hat{d}$	0.145	0.166	0.19	0.192	0.185	0.193	0.192	0.195	0.198	0.206	0.202	0.199

Table 7.17: Test for model $(1 - 2\nu B + B^2)^{0.1} = (1 + B)^{0.2} X_t = \varepsilon_t$ where ε_t is GARCH(1,1) and $\nu = -1$.

T	100	200	300	400	500	600	700	800	900	1000	2000	3000
n	20	43	57	65	82	81	79	91	84	89	100	100
$\hat{d}$	0.181	0.24	0.269	0.277	0.283	0.286	0.289	0.291	0.29	0.293	0.3	0.3

Table 7.18: Test for model $(1 - 2\nu B + B^2)^{0.3} X_t = \varepsilon_t$ where ε_t is GARCH(1,1), $\nu = \cos \frac{\pi}{3}$.

T	100	200	300	400	500	600	700	800	900	1000	2000	3000
n	2	8	17	18	39	39	51	52	50	63	85	100
$\hat{d}_1$	0.062	0.138	0.196	0.21	0.249	0.262	0.25	0.265	0.266	0.272	0.3	0.3
$\hat{d}_2$	0.218	0.27	0.331		0.341	0.362	0.368	0.367	0.374	0.383	0.394	0.4

Table 7.19: Test for model $(1 - 2\nu_1 B + B^2)^{0.3}(1 - 2\nu_2 B + B^2)^{0.4} X_t = \varepsilon_t$ where ε_t is GARCH(1,1), $\nu_1 = \cos \frac{\pi}{3}, \nu_2 = \cos \frac{5\pi}{6}$.

T	100	200	300	400	500	600	700	800	900	1000	2000	3000
n	0	5	10	10	23	41	38	51	62	58	95	97
$\hat{d}_1$	0.023	0.097	0.138	0.144	0.156	0.167	0.172	0.177	0.187	0.177	0.198	0.195
$\hat{d}_2$	0.167	0.224	0.249	0.266	0.281	0.285	0.277	0.281	0.284	0.283	0.302	0.3
$\hat{d}_3$	0.161	0.26	0.328	0.33	0.346	0.36	0.365	0.369	0.384	0.378	0.397	0.4

Table 7.20: Test for model $(1 - 2\nu_1 B + B^2)^{0.2}(1 - 2\nu_2 B + B^2)^{0.3}(1 - 2\nu_3 B + B^2)^{0.4} X_t = \varepsilon_t$ where ε_t is GARCH(1,1), $\nu_1 = \cos \frac{\pi}{6}, \nu_2 = \cos \frac{\pi}{2}, \nu_3 = \cos \frac{2\pi}{3}$.

Chapter 8

Conclusion

In this thesis, we have studied the stochastic long memory processes. Two classes of processes have been considered, the stationary long memory processes and the non-stationary long memory processes. We have investigated the probabilistic properties, parameter estimation methods, statistical tests, forecast methods, etc. And the applications to the financial data and the energy data are followed.

8.1 Overview of the Contribution

The contributions of this thesis on the stationary processes are composed by four parts.

- Self-similarity Properties:
  In Chapter 2, we studied the probabilistic properties of the stationary processes, focusing on their corresponding self-similar properties. In continuous-time framework, we have reviewed three classes of processes: H-self-similar processes (ex: Brownian motion), Gaussian H-self-similar process with stationary increments (fractional Brownian motion), non-Gaussian H-self-similar processes with stationary increments (ex: α-stable processes with $0 < \alpha < 2$) and multi-fractional processes. In discrete-time framework, we have reviewed the different definitions of self-similarity (exactly self-similar, asymptotically self-similar, exactly second-order self-similar, asymptotically second-order self-similar) and clarified the relationship among these definitions. Since the concepts of self-similarity and long / short memory are not equivalent, we proposed two important propositions which build the relationship between long range dependence and self-similarity. We proved that the stationary long memory process is asymptotically second-order self-similar, while the stationary short memory process is not asymptotically second-order self-similar. What's more, under Gaussianity, the stationary long memory process is asymptotically self-similar, and the stationary short memory process is not asymptotically self-similar. Then we applied these results on the classical discrete-time processes, for example, the fractional Gaussian noise, k-factor GARMA processes, heteroscedastic processes, processes with switches and jumps, etc.

- Estimation:
  In Chapter 4, we made a review of the estimation methods for stationary seasonal

and/or cyclical long memory (SCLM) processes (Robinson, 1994). Firstly we considered the ARFIMA models. After briefly describing the model, we recalled for this simple and classical long memory model the estimation methods, including the parametric, semiparametric (in frequency domain and in time domain) and wavelet methods. We did not present all the mathematical assumptions underlying the estimation procedure but rather described the methods and applicability, and we also discussed the asymptotic distributions of the estimators. For the parametric method, we have studied the four parametric maximum likelihood estimators (MLE): the exact time domain MLE, the modified profile likelihood estimator, the conditional time domain MLE and the frequency domain MLE. For the semiparametric estimators, we have looked at the log-periodogram regression method and the local Whittle Whittle approach. We also mentioned some modified versions of these estimators. For the wavelet methods, we focused on the wavelet-based ordinary least squares (OLS) estimator and the wavelet maximum likelihood estimator. We discussed also the advantages and disadvantages of these estimators. Secondly, we considered the estimation methods of the models with seasonality. We studied the k-factor Gegenbauer ARMA processes for its OLS estimator, approximate maximum likelihood estimator, conditional sum of squares residuals (CSS) estimator, wavelet-based OLS estimator and wavelet-based approximate maximum likelihood estimator, etc. We also have a brief look at the estimation methods for the models with fixed seasonal periodicity, the particular cases of the SCLM model. Thus, the rigid Hassler's models and flexible Hassler's models are all in this category. For completeness, we also mentioned a little the seasonal and/or cyclical asymmetric long memory processes and the corresponding estimation methods.

- Unit Root Test:
  In Chapter 7, we first recalled the well-known unit root test for ARMA process, Dicky-Fuller test, the augmented Dicky-Fuller test, etc. We mainly focused on one of unit root tests, the Robinson (1994) test. We evaluated the performances of the Robinson (1994) test for several simulated SCLM models. We showed that the sample size is crucial for the accuracy of the test. It seems that the use of this test is mainly recommended when we observe only one explosion in the spectral density if we have at least 500 points. The results raise concern as regards the applications of Robinson test to seasonal macroeconomics data.

Our contributions on the non-stationary processes are as follows:

- Modeling:
  In Chapter 5, we first recalled two kinds of non-stationary fractional integrated processes, the fractionally integrated processes with a constant long memory parameter $|d|(> \frac{1}{2})$ and the generalized fractionally integrated processes with time-varying parameter. Then we proposed a new non-stationary process, which permits the existence of the time-varying persistence and seasonality at the same time: the locally stationary k-factor Gegenbauer process. This new process can be regarded as an extension of the stationary k-factor Gegenabuer process allowing the time-varying

parameter. On the other hand, it can be considered as an extension of the generalized fractional integrated process permitting the seasonalities.

- Estimation and Forecast:
  In Chapter 5, we first reviewed the estimation methods for the non-stationary fractional integrated processes and their corresponding asymptotic behaviors.Then we proposed a new non-stationary model: the locally stationary k-factor Gegenabuer process and a new semiparametric estimation procedure using wavelet method, which is proved to be consistent and asymptotically normal. Through Monte Carlo simulation experiments, we studied the robustness of the estimator. The results are satisfying with small bias and small root mean square errors. Then we presented the forecast method for non-stationary long memory processes.

- Applications:
  Chapter 6 is concerned with the applications on the financial data and the energy data. For the error correction term in ECM of NAS 225 index, we applied on the series the new non-stationary model and the wavelet-based algorithm that we proposed in Chapter 5. With comparison to other authors' work, the estimation of our model captures better the local characteristics of the parameter function. And the forecast based on our estimation is also quite satisfying with little bias and RMSE. For the world crude oil data, we have modeled the WTI price series by several different models, the autoregressive models with fractional integrated noise and the locally stationary 1-factor Gegenabuer process. For the modeling by the locally stationary 1-factor Gegenabuer model, we need to pay attention to the range of the definition interval. If needed, we have to do some transformation such that the estimated parameter function lies well in the interval $(-\frac{1}{2}, \frac{1}{2})$. Thus, we can get the satisfying forecast results.

8.2 Possible Directions for Future Research

Some challenging problems remain for future research.

The first problem is to provide an approximate maximum likelihood estimation (AMLE) method for the locally stationary k-factor Gegenbauer processes. For the generalized fractional integrated model, generally, there exist two wavelet-based estimation methods: the wavelet-based OLS method and the wavelet-based AMLE method. In this thesis, we have built the wavelet-based OLS method for the locally stationary k-factor Gegenbauer processes. Thus, we would like to establish a wavelet-based AMLE method for this new non-stationary model. However, because of the technical difficulty, we have not yet solved this problem.

The second problem is to consider the locally stationary k-factor Gegenbauer processes with the short memory terms. What we consider in this thesis are the locally stationary k-factor Gegenabuer processes, which are the particular cases of the locally stationary k-factor GARMA processes. It will be of great interest to estimate the ARMA coefficients

together with the long memory parameter functions. But we are not quite sure whether the existence of ARMA term will result in some contaminations on the estimation.

Another idea is to propose a new model: the locally stationary seasonal and/or cyclical long memory process, which permits the long memory parameters of the SCLM model to vary with time. The spectrum will be a little complicated, and the wavelet-based estimation will be even more difficult. But it will be quite interesting to investigate such kind of general model.

Appendix A

The Well-definedness of the Locally Stationary k-factor Gegenbauer Processes

In this part, we give out the proof for Proposition 5.3.1. For more details and the classical theorem of Darboux's method, we can refer to Giraitis and Leipus (1995).

PROOF of Proposition 5.3.1.

Proof. We prove it through Darboux's method.
Let $\lambda_{-j} = -\lambda_j$, $d_{-j} = d_j$, $j = 1, \cdots, m$. Consider the following function

$$U^{d_1(t),\cdots,d_m(t)}(z) = \prod_{j=1}^{m}(1 - e^{i\lambda_j}z)^{d_j(t)}(1 - e^{-i\lambda_j}z)^{d_j(t)}$$

with singular points $z_j = e^{i\lambda_j}$, $j = \pm 1, \cdots, \pm m$ on the unit circle $|z| = 1$. Denote $I(k) = \{j : 1 \leq |j| \leq m, \lambda_j \neq \lambda_{-k} mod\, 2\pi\}$. Thus, $U^{d_1(t),\cdots,d_m(t)}(z) = e_k(t)h_k(1-zz_{-k})$ in the neighborhood of the point $z_k (1 \leq |k| \leq m)$, where $e_k(t) = \prod_{j\in I(k)} (1 - z_j z_k)^{d_j(t)}$ and $h_k(z,t) = z^{d_k^*(t)} \prod_{j\in I(k)} (1 - z_j z_k)^{d_k(t)}$. If we expand $h_k(z,t)$ in powers series about $z = 0$, we obtain $e_k(t)h_k(z,t) = \sum_{\nu=0}^{\infty} c_\nu^{(k)}(t) z^{d_k^*(t)+\nu}$, where

$$c_\nu^{(k)}(t) = e_k(t) \sum_{(s)} \prod_{j\in I(k)} \begin{pmatrix} d_j^*(t) \\ s_j \end{pmatrix} (\frac{z_j z_k}{1 - z_j z_k})^{s_j}$$

and the sum $\sum_{(s)}$ is taken over all integers $0 \leq s_j \leq \nu$, $j \in I(k)$ such that $\sum_{j\in I(k)} s_j = \nu$. Then by Darboux's method (see Giraitis and Leipus 1995), the following general

expansion for the weights $\pi_n(t)$ in (5.6) can be obtained

$$\pi_n(t) = \sum_{\nu=0}^{p-1} \sum_{k=1}^{m} \tilde{c}_\nu^{(k)} \begin{pmatrix} d_k^*(t) + \nu \\ n \end{pmatrix} + O(n^{-P-\min\{d_1^*(t),\cdots,d_m^*(t)\}-1}) \tag{A.1}$$

where

$$\tilde{c}_\nu^{(k)} = \begin{cases} 2Re(c_\nu^{(k)}(-e^{-i\lambda_k})^n), & if\ 0 < \lambda_k < \pi \\ Re(c_\nu^{(k)}(-e^{-i\lambda_k})^n), & \lambda_k = 0\ or\ \pi. \end{cases}$$

If we stop the expansion $(A.1)$ at the term $\nu = 0$ $(p = 1)$, we obtain

$$\pi_n(t) = \sum_{k=1}^{m} \begin{pmatrix} d_k^*(t) \\ n \end{pmatrix} \tilde{c}_0^{(k)} + O(n^{-2-\min\{d_1^*(t),\cdots,d_m^*(t)\}}),\ as\ n \to \infty.$$

Then the symmetry leads to the asymptotical expansion (5.9). So the application of the linear filter $(\psi_j,\ j \in \mathbb{Z})$ to a stationary white noise $\varepsilon(t)$ gives a well-defined process

$$\tilde{X}(t) = \bigtriangledown_{\lambda_1,\cdots,\lambda_m}^{-d_1(t),\cdots,-d_m(t)} = \sum_{n=0}^{\infty} \psi_n(t)\varepsilon(t-n),$$

which is the solution of Equation (5.13).

According to the condition (5.7) and asymptotical expansion (5.9), we can get $\sum_{n=0}^{\infty} \psi_n^2(t) < \infty$ and $\sum_{n=0}^{\infty} \pi_n^2(t) < \infty$. Therefore we obtain the uniqueness. □

Bibliography

[1] Abadir, K.M., W. Distaso, L. Giraitis (2007). Nonstationarity-extended local Whittle estimation. *J. Econometrics*, 141, 2, 1353-1384.

[2] Abry, P., P. Flandrin, M.S. Taqqu and D. Veitch (2000). Wavelets for the analysis, estimation ans synthesis of scaling data. In: K.Park and W. Willinger (eds.), Self-similar Network Traffic and Performance Evaluation. Wiley, New York.

[3] Abry, P., P. Flandrin, M.S. Taqqu and D. Veitch (2001). Self-similarity and long-range dependence through the wavelet lens. In: P. Doukhan, G. Oppenheim and M.S. Taqqu, eds., Long-Range dependence: Theory and Applications. Birkhäuser, Boston.

[4] Abry, P. and F. Sellan (1996). The wavelet based synthesis for fractional Brownian motion proposed by F. Sellan and Y. Meyer: Remarks and fast implementation. *Applied and Computational Harmonic Analysis,* 3, 377-383.

[5] Agiakloglou, C., P. Newbold and M. Wohar (1993). Bias in an estimator of the fractional difference parameter. *J. Time Series Analysis*, 14, 235-246.

[6] An, S. and P. Bloomfield (1993). Cox and Reid's modification in regression models with correlated errors. Technical Report, Department of Statistics, North Carolina State University, Raleigh, NC 27695-8203, USA.

[7] Anděl, J. (1986). Long memory time series models. *Kybernetika*, 22, 105-123.

[8] Andersen, T.G., T. Bollerslev, F.X. Diebold and H. Ebens (2001). The distribution of realized stock return volatility. *J. Financial Economics*, 61, 1, 43-76.

[9] Andersen, T.G., T. Bollerslev, F.X. Diebold and P. Labys. The Distribution of Realized Exchange Rate Volatility. *J. American Statistical Association*, 96(453), 42-55.

[10] Andrews, D.W.K. and P. Guggenberger (2003). A bias-reduced log-periodogram regression estimator for the long-memory parameter. *Econometrica*, 71, 2, 675-712.

[11] Andrews, D.W.K. and Y. Sun (2004). Adaptive Local Polynomial Whittle Estimation of Long-Range Dependence. *Econometrica*, 72, 2, 569-614.

[12] Arteche, J. (1998). Log-periodogram regression in seasonal/cyclical long memory time series. Working paper, university of the Basque country (UPV-EHU), November.

[13] Arteche, J. and P.M. Robinson (2000). Semiparametric inference in seasonal and cyclical long memory processes. *J. Time Series Analysis*, 21, 1-25.

[14] Arteche, J. (2003). Semi-parametric robust tests on seasonal or cyclical long memory time series. *J. Time Series Analysis*, 23, 251-285.

[15] Baillie R.T. (1996). Long memory processes and fractional integration in econometrics. *J. Econometrics*,73, 1, 5-59.

[16] Baillie, R.T., T. Bollerslev and H.O. Mikkelsen (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *J. Econometrics,* 74, 3-30.

[17] Baillie, R.T. and H. Chung (2001). Estimation of GARCH models from the autocorrelations of the squares of a process. *J. Time Series Analysis,* 22, 631-650.

[18] Bardet, J.M., G. Lang, E. Moulines and P. Soulier (2000). Wavelet Estimator of long-range dependent processes. *Statistical Inference for Stochastic Processes,* 3, 85-99.

[19] Bardet, J.M., G. Lang, G. Oppenheim, A. Philippe, S. Stoev and M.S. Taqqu (2001). Semi-parametric estimation of the long-range dependence parameter: A survey. In: P. Doukhan, G. Oppenheim and M.S. Taqqu (eds.). Long-Range Dependence: Theory and Applicaitons. Birkhäuser, Boston.

[20] Barkoulas, J.T. and C.F. Baum (1997). Fractional differencing modeling and forecasting of eurocurrency. *J. Financial Research*, 20, 3, 355-372.

[21] Beran, J. (1994). Statistics for long memory processes. Chapman and Hall, New York.

[22] Beran, J. (1995). Maximum likelihood estimation of the differencing parameter for invertible short and long memory autoregressive integrated moving average models. *J. Royal Statistical Society*, 57, 659-672.

[23] Beran, J. and N. Terrin (1996). Testing for a change of the long-memory parameter. *Biometrika*, 83 (3), 627-638.

[24] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *J. Econometrics*, 31, 307-327.

[25] Bollerslev, T. (1988). On the corrrelation structure for the generalized autoregressive conditonal heteroskedastic process. *J. Time Series Analysis*, 9, 121-131.

[26] Briggs, W.L. and V.E. Henson (1995). The DFT: an owner's manual for the discrete Fourier transform. Society for Industrial and Applied Mathematics Philadelphia.

[27] Brockwell, P.J. and R.A. Davis (2002). Introduction to Time Series and Forecasting. Series: Springer Texts in Statistics, Springer.

[28] Brodsky, J. and C.M. Hurvich (1999). Multi-step forecasting for long-memory processes. *J. Forecasting*, 18, 1, 59-75.

[29] Burrus, C.S., R.A. Gopinath and H. Guo (1998). Introduction to wavelets and wavelet transforms. Prentice Hall Upper Saddle River, NJ.

[30] Carlin, J.B. and A.P. Dempster (1989). Sensitivity analysis of seasonal adjustements: Empirical cas studies. *J. American Statistical Association*, 84, 6-20.

[31] Cavanaugh, J.E., Y. Wang and J.W. Davis (2002). Locally self-similar processes and their wavelet analysis. Chapter 3 in Handbook of statistics 21: Stochastic processes: Modelling and simulation, DN Shanbhag and CR Rao (eds.). Elsevier Science, Amsterdam, The Netherlands.

[32] Chan, G. and A.T.A. Wood (1994). Simulation of stationary Gaussian processes in $[0,1]^d$. *J. Compututational and Graphical Statistics*, 3, 409-432.

[33] Chen, G., P. Hall and D.S. Poskitt (1995). Periodogram-based estimators of fractal properties. *The Annals of Statistics*, 23, 1684-1711.

[34] Cheung, Y.W. and F.X. Diebold (1994). On maximum-likelihood estimation of the differencing parameter of fractionally-integrated noise with unknown mean. *J. Econometrics*, 62, 301-316.

[35] Cheung, Y.W. and K.S. Lai (1993). A search for long memory in international stock market returns. *J. International Money and Finance*, 14, 4, 597-615.

[36] Chui, C. K. (1992). An Introduction to Wavelets. New York: Academic Press.

[37] Chui, C. K. (1997). Wavelets: A Mathematical Tool for Signal Analysis. Society for Industrial Mathematics.

[38] Chung, C.F. (1996a). Estimating a generalized long memory process. *J. Econometrics*, 73, 1, 237-259.

[39] Chung, C.F. (1996b). A generalized fractionally integrated autoregressive moving-average process. *J. Time Series Analysis*, 17, 2, 111-140.

[40] Chung, C.F. and R.T. Baillie (1993). Small sample bias in conditional sum-of-squares estimators of fractionally integrated ARMA models. *Empirical Economics*, 18, 791-806.

[41] Coifman, R.R. and D. Donoho (1995). Time-invariant wavelet de-noising. *In Antoniadis and Oppenheim*, 125-150.

[42] Collet, J. and D. Guégan (2004). Another Characterization of the long memory behavior. Note de Recherche IDHE-MORA No01-2004, ENS Cachan, France.

[43] Comte, F. (1996). Simulation and estimation of long memory continous time models. *J. Time Series Analysis,* 17, 19-36.

[44] Constantine, A.G. and P. Hall (1994). Characterizing surface smoothness via estimation of effective fractal dimension. *J. Royal Statistical Society B*, 56, 97-113.

[45] Cooley, J.W. and J.W. Tukey (1965). An Algorithm for the Machine Calculation of Complex Fourier Series. *Mathematics of computation*,19, 90, 297-301.

[46] Cooley, J.W., P.A.W. Lewis and P.D. Welch (1967). Historical notes on the fast Fourier transform. *Proceedings of the IEEE*, 55, 10, 1675-1677.

[47] Cox, D.R. (1984). Long-range dependence: A review, Statistics: An Appraisal. Proceedings 50th Anniversary Conference, Iowa State Statistical Laboratory, Iowa State University Press, 55-74.

[48] Cox, D.R. and N. Reid (1987). Approximations to Noncentral Distributions. *The Canadian Journal of Statistics*, 15, 2.

[49] Cox, D.R. and N. Reid (1992). A note on the difference between profile and modified profile likelihood. *Biometrika*, 79(2), 408-411.

[50] Cox, D.R. and N. Reid (1993). A note on the calculation of adjusted profile likelihood. *J. Royal Statistical Society*, Series B.

[51] Crato, N. and B.K. Ray (1999). Model selection and forecasting for long-range dependent processes. *J. Forecasting*,15, 107-125.

[52] Dahlhaus, R. (1989). Efficient parameter estimation for self-similar processes. it Annals of Statistics, 17, 1749-1766.

[53] Dahlhaus, R. (1996a). On the Kullback Leibler information divergence of locally stationary processes. *Stochastic Processes and their Applications*, 62, 139-168.

[54] Dahlhaus, R. (1996b). Asymptotic statistical inference for nonstationary processes with evolutionary spectra. In Athens Conference on Applied Probability and Time Series Analy Z. sis P. M. Robinson and M. Rosenblatt (eds.). 2. Springer, New York.

[55] Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *Annals of Statistics*, 25, 1, 37.

[56] Darné, O., V. Guiraud and M. Terraza (2004). Forecast of the seasonal fractional integrated series. *J. Forecasting*, 23, 1-17.

[57] Daubechies, I. (1988). Orthonormal bases of compactly supported wavelets. *Communications on Pure and Applied Mathematics*, 41, 909-996.

[58] Daubechies, I. (1992). Ten Lectures on Wavelets. Volume 61 of CBMS-NSF Regional Conference Series in Applied Mathematics. Philadelphia: Society for Industrial and Applied Mathematics.

[59] Davidson, J. (2001). Moment and memory properties of linear conditional heteroscedasticity models. Working paper, Cardiff University, UK.

[60] Davies, R.B. and D.S. Harte (1987). Tests for Hurst effect. *Biometrika*, 74(1), 95-101.

[61] Dickey, D.A. (1976). Estimation and hypothesis testing in nonstationary time series. PHD dissertation, Iowa State University.

[62] Dickey, D.A. and W.A. Fuller (1979). Distribution of the Estimator for Autoregressive Time Series with a Unit Root. *J. American Statistical Association*, 74, 427-431.

[63] Dickey, D.A. and W.A. Fuller (1981). Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root. *Econometrica*, 49, 1057-1072.

[64] Dickey, D.A., W.R. Bell and R.B. Miller (1986). Unit roots in time series models: tests and implications. *The American Statistician*, 40, 1, 12-26.

[65] Diebold, F.X. and A. Inoue (2001). Long memory and regime switching. *J. Econometrics*, 105, 1, 131-159.

[66] Diebold, F.X., S. Husted and M. Rush (1991). Real Exchange Rates under the Gold Standard. *The Journal of Political Economy*, 99, 6, 1252-1271.

[67] Diebold, F.X. and G. D. Rudebusch (1989). Long memory and persistence in aggregate output. *J. Monetary Economics*, 24(2), 189-209.

[68] Diebold, F.X. and G. D. Rudebusch (1991). Forecasting output with the composite leading index: a real-time analysis. *American Statistical Association*, 86, 415.

[69] Ding, Z. and C.W.J. Granger (1996). Modeling volatility persistence of speculative returns: A new approach. *J. Econometrics,* 73, 185-215.

[70] Diongue, A.K., D. Guégan and B. Vignal (2007). The stationary seasonal hyperbolic asymmetric power ARCH model. *statistics and probability letters*, 77, 1158-1169.

[71] Diongue, A.K., D. Guégan and B. Vignal (2009). Forecasting electricity spot market prices with a k-factor GIGARCH process. *Applied energy*, 36, 505-510.

[72] Diongue, A.K., D. Guégan and B. Vignal (2004). A k-factor GIGARCH process: estimation and application on electricity market spot prices. IEEE proceedings of the 8th International Conferences on probability methods Applied to power systems, Iowa State University, AMES, Iowa, 1 - 7.

[73] Dolado, J.J., J. Gonzalo and L. Mayoral (2002). A fractional Dickey-Fuller test for unit roots. *Econometrica*, 70, 5, 1963-2006.

[74] Dufrénot, G., D. Guégan and A. Péguin-Feissolle (2005a). Long-memory dynamics in a SETAR model. Applications to stock markets. *J. International Financial Markets, Institutions and Money*, 15, 391-406.

[75] Dufrénot, G., D. Guégan and A. Péguin-Feissolle (2005b). Modeling squared returns using a SETAR model with long-memory dynamics. *Economics Letters*, 86, 237-243.

[76] Dufrénot, G., D. Guégan and A. Péguin-Feissolle (2008). Changing regime volatility: a fractionally integrated SETAR model. *Applied Financial Economics*, 18, 519-526.

[77] Embrechts, P. and M. Maejima (2000). An introduction to the theory of self-similar stochastic processes. *International Journal of Modern Physics B,* 14, 1399-1420.

[78] Engle, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kindom inflation. *Econometrica,* 50, 987-1007.

[79] Engle, R.F. and T. Bollerslev (1986). Modeling the persistence of conditional variance, *Econometric Reviews*, 5, 1-50.

[80] Engle, R.F. and C.W.J. Granger (1987). Co-integration and error correction: Representation, estimation and testing. *Economica*, 55, 251-276.

[81] Engle, R.F. and B.S. Yoo (1987). Forecasting and testing in cointegrated systems. *J. Econometrics*, 35, 143-159.

[82] Evans, G.B.A. and N.E. SAVIN (1981). Testing for Unit Roots: 1. *Econometrica*, 49, 753-779.

[83] Evans, G.B.A. and N.E. SAVIN (1984). Testing for Unit Roots: 2. *Econometrica*, 52, 1241-1269.

[84] Fan, J. (2000). Prospects of nonparametric modeling. *J. American Statistical Association*, 95, 452, 1296-1300.

[85] Fan, J. and Q. Yao (2005). Nonlinear time series: Nonparametric and parametric methods. Springer Science, Business Media, Inc.

[86] Ferrara, L. (2000). Processus longue méoire généralisés: estimation, prévision et applications. thèse de doctorat, Université Paris XIII, France.

[87] Ferrara, L. and D. Guégan (2001a). Forecasting with k-factor Gegenbauer processes: Theory and applications. *J. Forecasting*, 20, 581-601.

[88] Ferrara, L. and D. Guégan (2001b). Comparison of parameter estimation methods in cyclical long memory time series. Developments in Forecast Combination and Portfolio Choice, C. Dunis, J. Moody and A. Timmermann (eds.), Wiley, New York, Chapter 8.

[89] Fox, R. and M.S. Taqqu (1986). Large-sample properties of parameter estimates for strongly dependent stationary gaussian series. *Annals of Statistics*, 14, 517-532.

[90] Franses, P.H. and Ooms, M. (1997). A periodic long memory model for quartely UK inflation. *International Journal of Forecasting*, 13, 117 - 126.

[91] Fuller, W.A. (1976). Introduction to statistical time series. New York, John Wiley and Sons.

[92] Gençay, R., B. Selçuk and B. Whitcher (2001). An introduction to wavelets and other filtering methods in finance and economics(1 ed.). Academic Press.

[93] Geweke J. and S. Porter-Hudak (1983). The estimation and application of long memory time series models. *J. Time Series Analysis*, 4, 221-238.

[94] Gil-Alana, L.A. (2001). A fractionally integrated exponential spectral model for the UK unemployement. *J. Forecasting*, 20, 329-340.

[95] Gil-Alana, L.A. (2006). Testing seasonality in the contrext of fractionally integrated processes. *Annales d'Economie et de Statistique*, 81, 69-91.

[96] Gil-Alana, L.A. and P.M. Robinson (1997). Testing of unit root and other nonstationary hypotheses in macroeconomic time series. *J. Econometrics*, 80, 2, 241-268.

[97] Gil-Alana, L.A. and P.M. Robinson (2001). Testing of seasonal fractional integration in UK and Japanese consumption and income. *J. Applied Econometrics*, 16(2), 95-114.

[98] Giraitis, L. and R. Leipus (1995). A generalized fractionally differencing approach in long memory modelling. *Lithuanian Mathematical Journal*, 35, 65-81.

[99] Giraitis, L. and P.M. Robinson (2003). Edgeworth Expansions for Semiparametric Whittle Estimation of Long Memory. *The Annals of Statistics*, 31, 4, 1325-1375.

[100] Giraitis, L. and D. Surgailis (1990). A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotic normality of Whittleąŕs estimate. *Probability Theory and Related Fields*, 86, 87-104.

[101] Goncalves, P. and P. Abry (1997). Multiple-window wavelet transform and local scaling exponent estimation. *Acoustics, Speech, and Signal Processing, 1997. ICASSP-97., 1997 IEEE International Conference on*, 5, 3433-3436.

[102] Goupillaud. P., A. Grossmann and J. Morlet (1984). Cycle-octave and related transforms in seismic signal analysis. *Geoexploration*, 23,1, 85-102.

[103] Granger, C.W.J. and R. Joyeux (1980). An introduction to long memory time series models and fractional differencing. *J. Time Series Analysis*, 1, 15-29.

[104] Granger, C.W.J. and T. Teräsvirta (1999). A simple nonlinear time series model with misleading linear properties. *Economics letters,* 62, 161-165.

[105] Gray, H.L., N.F. Zhang and W.A. Woodward (1989). On generalized fractional processes. *J. Time Series Analysis*, 10(3), 233-257.

[106] Gray, H.L., N.F. Zhang and W.A. Woodward (1994). On generalized fractional processes - a correction. *J. Time Series Analysis*, 15(5), 561-562.

[107] Green, P.J. and B.W. Silverman (1994). Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach. Chapman and Hall.

[108] Grossmann, A. and J. Morlet (1984). Decomposition of functions into wavelets of constant shape, and related transforms. University of Bielefeld.

[109] Guégan, D. (2000). A new model: the k-factor GIGARCH process. *J. Signal processing,* 4, 265-271.

[110] Guégan, D. (2003). A prospective study of the k-factor Gegenbauer process with heteroscedastic errors and an application to inflation rates. *Finance India*, 17, 1-21.

[111] Guégan, D. (2005). How can we define the concept of long memory? An econometric survey. *Econometric Reviews,* 24, 2.

[112] Guégan, D. (2009). Global and local stationary modelling in finance: theory and empirical evidence. In revision for Economic Review.

[113] Guégan, D. and Z.P. Lu (2007). A note on self-similarity for discrete time series. CES Working paper, 2007.55, Université Paris 1, France.

[114] Hall, P., H.L. Koul and B.A. Turlach (1997). Note on convergence rates of semiparametric estimators of dependence index. *The Annals of Statistics,* 25, 1725-1739.

[115] Haar, A. (1910). Zur Theorie der orthogonalen Funktionensysteme. *(German) Mathematische Annalen*, 69, 3, 331-371.

[116] Hassler, U. (1994). Misspecification of long memory seasonal time series. *J. Time Series Analysis*, 15, 19-30.

[117] Hassler, U. and J. Wolters (1994). On the power of unit root test against fractional alternatives. *Economics Letters*, 45, 1-5.

[118] Hauser, M.A., B.M. Potscher, E. Reschenhofer (1999). Measuring persistence in aggregate output: ARMA models, fractionally integrated ARMA models and nonparametric procedures. *Empirical Economics*, 24, 2, 243-269.

[119] Henry, M. and P.M. Robinson (1996). Bandwidth choice in Gaussian semiparametric estimation of long range dependence, in P. M. Robinson and M. Rosenblatt (eds.). *Athens Conference on Applied Probability and Time Series Analysis, Volume II: Time Series Analysis, In Memory of E. J. Hannan*, Springer, New York, 220-232.

[120] Hipel, K.W. and A.I. McLeod (1978). Preservation of the rescaled adjusted range, Part 2, Simulation studies using Box-Jenkins models. *Water Resources Research*, 14, 509-516.

[121] Hosking, J.R.M. (1981). Fractional differencing. *Biometrika*, 68(1), 165-176.

[122] Hosoya, Y. (1997). A limit theory for long-range dependence and statistical inference on related models. *Annals of Statistics*, 25, 105-137.

[123] Hubbard, B.B. (1996). The world according to wavelets. AK Peters, Ltd., 2 Rev Upd edition.

[124] Hurst, H.E. (1951). Long term storage capacity of reservoirs, *American Society of Civil Engineers*, 116, 776-808.

[125] Hurvich, C.M. and K.I. Beltrao (1993). Asymptotics for the low-frequency ordinates of the periodogram of a long-memory time series. *J. Time Series Analysis*, 14, 5, 455-472.

[126] Hurvich, C.M. and K.I. Beltrao (1994). Automatic Semiparametric Estimation of the Memory Parameter of a Long-Memory Time Series. *J. Time Series Analysis*, 15, 285-302.

[127] Hurvich, C.M. and W.W. Chen (2000). An efficient taper for potentially overdifferenced longmemory time series. *J. Time Series Analysis*, 21, 155-180.

[128] Hurvich, C.M., R.S. Deo and J. Brodsky (1998). The mean squared error of Geweke and Porter-Hudakąŕs estimator of the memory paramater of a long memory time series. *J. Time Series Analysis*, 19, 19-46.

[129] Hurvich, C.M. and B.K. Ray (1995). Estimation of the memory parameter for nonstationary or noninvertible fractionally integrated processes. *J. Time Series Analysis*, 16, 17-42.

[130] Jensen, M.J. (1999a). An approximate wavelet MLE of short and long memory parameters. *Studies in Nonlinear Dynamics and Economics*, 3(4), 239-253.

[131] Jensen, M.J. (1999b). Using wavelets to obtain a consistent ordinary least squares estimator of the long-memory parameter. *J. Forecasting*, 18(1), 17-32.

[132] Jensen, M.J. (2000). An alternative maximum likelihood estimator of long-memory processes using compactly supported wavelets. *J. Economic Dynamics and Control*, 24, 3, 361-387.

[133] Jensen, M.J. and B. Whitcher (2000). Time-varying long-memory in volatility: detection and estimation with wavelets. Working Paper, Department of Economics, University of Missouri, Columbia.

[134] Jones, R.H. (1971). Spectrum estimation with missing observations. *Annals of the Institute of Statistical Mathematics*, 23, 1, 387-398.

[135] Johnstone, I.M. and B.W. Silverman (1997). Wavelet threshold estimators for data with correlated noise. *J. Royal Statistical Society. Series B*, 59, 319-351.

[136] Karanasos, M., Z. Psaradakis and M. Sola (2004). On the autocorrelation properties of long-memory GARCH processes. *J. Time Series Analysis*, 25, 265-281.

[137] Kent, J.T. and A.T.A. Wood (1997). Estimating the fractal dimension of a locally self-similar Gaussian process by using increments. *J. Royal Statistical Society B,* 59, 679-700.

[138] Kim, C.S. and P.C.B. Phillips (1999). Modified log periodogram regression. mimeo, Yale University.

[139] Kolmogorov, A.N. (1941). On degeneration of isotropic turbulence in an incompressible viscous liquid. Dokl. Akad. Nauk SSSR, 30, 301.

[140] Kolmogorov, A.N. (1961). Local structure of turbulence in fluid for every large Reynolds numbers, Transl. in Turbulence. S.K.Friedlander and L.Topper (eds). Interscience Publishers, New York, 151-155.

[141] Kolmogorov, A.N. (1991). The Local Structure of Turbulence in Incompressible Viscous Fluid for Very Large Reynolds Numbers. *Proceedings: Mathematical and Physical Sciences*, 434, 1890, Turbulence and Stochastic Process: Kolmogorov's Ideas 50 Years On (Jul. 8, 1991), 9-13.

[142] Kozhemyak, A. (2006). Modélisation de séries financières à l'aide de processus invariants d'échelle. Application à la prédiction du risque, thèse de doctorat, École Polytechnique, France.

[143] Krämer, W. and I. Dittmann (1998). Fractional integration and the augmented Dickey-Fuller Test. *Economics Letters*, 61, 269-272.

[144] Künsch, H.R. (1986). Discrimination between monotonic trends and long-range dependence. *J. Applied Probability*, 23, 1025-1030.

[145] Künsch, H.R. (1987). Statistical aspects of self-similar processes. In Y. Prokhorov and V.V. Sazanov (eds.). *Proceedings of the First World Congress of the Bernoulli Society*, VNU Science Press, Utrecht, 67-74.

[146] Leland, W.E., M.S. Taqqu, W. Willinger and D.V. Wilson (1994). On the self-similar nature of Ethernet traffic (extended version). *IEEE/ACM Trans Netw.*, 2, 1(Feb.), 1-15.

[147] Liang, J. and T.W. Parks (1996). A translation-invariant wavelet representation algorithm with applications. *IEEE Transactions on Signal Processing*, 44, 2, 225-232.

[148] Lien, D. and Y.K. Tse (1999). Forecasting the Nikkei spot index with fractional cointegraion. *J. Forecasting*, 18, 259-273.

[149] Lo, A.W. (1991). Long term memory in stock market prices. *Econometrica*, 59, 1279-1313.

[150] Lobato, I. and P.M. Robinson (1996). Averaged periodogram estimation of long memory. *J. Econometrics*, 73, 1, 303-324.

[151] López-Ardao, J.C., C. López-García, A. Suárez-González, M. Fernández-Veiga and R.F. Rodríguez-Rubio (2000). On the use of self-similar processes in network simulation. *ACM Transactions on Modeling and Computer Simulation*, 10, 125-151.

[152] Mallat, S. (1989). A theory for multiresolution signal decomposition: the wavelet-trepresentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions*, 11, 7, 674-693.

[153] Mallat, S. (1999). A wavelet tour of signal processing (2 ed.). Academic press.

[154] Mallat, S. and W.L. Hwang (1992). Singularity detection and processing with wavelets. *IEEE Transactions on Information Theory*, 38, 2.

[155] Mallat, S. and D. Zhong (1992). Characterization of signals from multiscale edges. *IEEE Transactions on patiern analysis and machine intelligence*, 14, 7.

[156] Mandelbrot, B.B. and J.W. Van Ness (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Review*, 10, 422-437.

[157] Mandelbrot, B.B. and J.R. Wallis (1969). Computer experiments with fractional Gaussian noises. *Water Resources Resarch*, 5, 228-267.

[158] Maynard, A. and P.C.B. Phillips (2001). Rethinking an old empirical puzzle: econometric evidence on the forward discount anomaly. *J. Applied Econometrics*, 16, 6, 671-708.

[159] McCoy, E.J. and D.A. Stephens (2004). Bayesian time series analysis of periodic behaviour and spectral structure. *International Journal of Forecasting*, 20, 4, 713-730.

[160] McCoy, E.J. and A.T. Walden (1996). Wavelet analysis and synthesis of stationary long-memory processes. *J. Computational and Graphical Statistics*, 5(1), 26-56.

[161] Meyer, Y. (1991). Wavelets and applications. *Proceedings of the International Congress of Mathematicians*, I, II, 1619-1626.

[162] Meyer, Y. (1992). Wavelets and operators. Cambridge Studies in Advanced Math., vol. 37, Cambridge Univ. Press, Cambridge.

[163] Morris, J.M. and R. Peravali (1999). Minimum-bandwidth discrete-time wavelets. *Signal Processing*, 76(2), 181-193.

[164] Moulines, E., F. Roueff and M.S. Taqqu (2008). A wavelet whittle estimator of the memory parameter of a nonstationary gaussian time series. *The Annals of Statistics*, 36, 4, 1925-1956.

[165] Nason, G.P. and B.W. Silverman (1995). The stationary wavelet transform and some statistical applications. *In Antoniadis and Oppenheim*, 281-300.

[166] Nelson, D.B. (1990). Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory*, 6, 318-334.

[167] Nielsen, M.Ø. (2004), Efficient likelihood inference in nonstationary univariate models. *Econometric Theory*, 20, 116-146.

[168] Nielsen, M.Ø. and P.H. Frederiksen (2005). Finite sample comparison of parametric, semiparametric, and wavelet estimators of fractional integration. *Econometric Reviews*, 24, 405-443.

[169] Noakes, D.J., K.W. Hipel, A.I. McLeod, C. Jimenez and C. Yakowitz (1988). Forecasting annual geophysical time series. *International Journal of Forecasting*, 4, 103-115.

[170] Ogden, R.T. (1997). Essential Wavelets for Statistical Applications and Data Analysis. Boston: Birkhäuser.

[171] Olhede, S.C., E.J. McCoy and D.A. Stephens (2004). Large-sample properties of the periodogram estimator of seasonally persistent processes. *Biometrika*, 91(3), 613-628.

[172] Palma, W. (2000). Long-Memory Time Series: Theory and Methods. Wiley Series in Probability and Statistics.

[173] Palma, W. and N.H. Chan (2005). Efficient Estimation of Seasonal Long-Range-Dependent Processes. *J. Time Series Analysis*, 26, 6, 863-892.

[174] Palma, W. and M. Zevallos (2004). Analysis of the correlation structure of square time series. *J. Time Series*, 25, 529-550.

[175] Parke, W.R. (1999). What is Fractional Integration? *The review of economics and statistics*, 81, 4, 632-638.

[176] Percival, D.B. (1993). Simulating Gaussian random processes with specified spectra. *Computing Science and Statistics*, 24, 534-538.

[177] Percival, D.B. (1995). On estimation of the wavelet variance. *Biometrika*, 82(3), 619-631.

[178] Percival, D.B. and H.O. Mofjeld (1997). Analysis of Subtidal Coastal Sea Level Fluctuations Using Wavelets. *J. American Statistical Asscociation*, 92, 868-880.

[179] Percival, D.B. and A.T. Walden (2000). Wavelet methods for time series analysis. Cambridge University Press.

[180] Pesquet, J.C., H. Krim and H. Carfantan (1996). Time-invariant orthonormal wavelet representations. *IEEE Transactions on Signal Processing*, 44, 8, 1964-1970.

[181] Phillips, P.C.B. (1987). Time Series Regression with a Unit Root. *Econometrica*, 55, 277-301.

[182] Phillips, P.C.B. (2007). Unit root log periodogram regression. *J. Econometrics*, 138, 1, 104-124.

[183] Phillips, P.C.B. and K. Shimotsu (2000). Modified Local Whittle Estimation of the Memory Parameter in the Nonstationary Case. Cowles Foundation Discussion Papers of Yale Universitywith number 1265.

[184] Phillips, P.C.B. and K. Shimotsu (2004). Local Whittle estimation in nonstationary and unit root cases. *Annals of Statistics*, 32, 656-692.

[185] Phillips, P.C.B. and K. Shimotsu (2006). Local Whittle estimation of fractional integration and some of its variants. *J. Econometrics*, 130, 2, 209-233.

[186] Porter-Hudak, S. (1990). An application to the seasonally fractionally differenced model to the monetary aggregates. *J. American Statistical Association*, 85, 338-344.

[187] Priestley, M.B. (1981). Spectral Analysis and Time Series. Academic Press, London.

[188] Rainville, E.D. (1960). Special functions. The Macmillan Company, New York.

[189] Ramsey J.B. (1998). Regression over time scale decomposition: a sampling analysis of distribitional properties. *Economic Systems Research*, 11, 163-183.

[190] Ramsey, J.B. (1999). The contribution of wavelets to the analysis of economic and financial data. Philosophical Transactions of the Royal Society.

[191] Ramsey, J.B. (2002). Wavelets in Economics and Finance: Past and Future. *Studies in Nonlinear Dynamics and Econometrics*, 6, 3, 1.

[192] Ramsey, J.B. and C. Lampart (1998a). The Decomposition of Economic Relationships by Time Scale Using Wavelets: Expenditure and Income. *Studies in Nonlinear Dynamics and Econometrics*, 3, 1, 2.

[193] Ramsey, J.B. and C. Lampart (1998b). Decomposition of economic relationships by timescale using wavelets. *Macroeconomic Dynamics*, 2, 49-71.

[194] Ramsey, J.B. and Z. Zhang (1997). The analysis of foreign exchange data using waveform dictionaries. *J. Empirical Finance*, 4, 4, 341-372.

[195] Ray, B.K. (1993a). Modelling long memory processes for optimal long range prediction. *J. Time Series Analysis*, 14, 511-526.

[196] Ray, B.K. (1993b). Long-range forecasting of IBM product revenues using a seasonal fractionally differenced ARMA model. *International Journal of Forecasting*, 9, 255-269.

[197] Reisen, V.A. (1994). Estimation of the fractional difference parameter in the ARFIMA(p,d,q) model using the smoothed periodogram. *J. Time Series Analysis*, 15(1), 335-350.

[198] Robinson, P.M. (1976). Instrumental Variables Estimation of Differential Equations. *Economitrica*, 44, 4, 765-776.

[199] Robinson, P.M. (1992). Log-periodogram regression for time series with long range dependence. Unpublished manuscript.

[200] Robinson, P.M. (1994). Efficient tests of nonstationary hypotheses. *J. American Statistical Association*, 89(428), 1420-1437.

[201] Robinson, P.M. (1995). Log-periodogram regression of time series with long range dependence. *Annals of Statistics*, 23, 1048-1072.

[202] Said, E.S. and D.A. Dickey (1984). Testing for unit roots in ARMA(p,q) models with unknown p and q. *Biometrika*, 71, 599-607.

[203] Samorodnitsky, G. and M.S. Taqqu (1994). Stable non-Gaussian random Processes stochastic models with infinite variance. Chapman and Hall, New York.

[204] Sargan, J.D. and A. Bhargava (1983). Testing Residuals from Least Squares Regression for Being Generated By a Gaussian Random Walk. *Econometrica* 51, 153-174.

[205] Schmidt, P. and J. Lee (1996). A modification of the Schmidt-Phillips unit root test. *Economics Letters*, 36, 3, 285-289.

[206] Shann, W.C. and C.C. Yen (1999). On the exact values of orthonormal scaling coefficients of lengths 8 and 10. *Applied and Computational Harmonic Analysis*, 6, 1, 109-112.

[207] Shimotsu, K. (2002). Exact local Whittle estimation of fractional integration with unknown mean and time trend. Department of Economics Discussion Paper No. 543, University of Essex.

[208] Shimotsu, K. and P.C. B. Phillips (2002). Pooled log periodogram regression. *J. Time Series Analysis*, 23, 57-93.

[209] Shimotsu, K. and P.C.B. Phillips (2005). Exact local Whittle estimation of fractional integration. *The Annals of Statistics*, 33, 4, 1890-1933.

[210] Sibbertsen, P. (2004). Long memory versus structural breaks: An overview. *Statistical Papers*, 45, 4.

[211] Smith, J. and S. Yadav (1994). Forecasting cost incurred from unit differencing fractionally integrated processes. *International Journal of Forecasting*, 10, 507-514.

[212] Sowell, F. (1989). A decomposition of block toeplitz matrices with applications to vector time series. Technical Report, GSIA, Carnegie Mellon University.

[213] Sowell, F. (1992a). Modeling long-run behavior with the fractional ARIMA model. *J. Monetary Economics*, 29(2), 277-302,

[214] Sowell, F. (1992b). Maximum likelihood estimation of stationary univariate fractionally integrated time series models. *J. Economics*, 53, 165-188.

[215] Strang, G. and T. Nguyen (1996). Wavelets and filter banks. Wellesley-Cambridge Press, Wellesley, MA.

[216] Stengos, T. and Y. Sun (2001). A consistent model specification test for a regression function based on nonparametric wavelet estimation. *Econometric Reviews*, 20, 1, 41-60.

[217] Tanaka, K. (1999). The nonstationary fractional unit root. *Econometric theory*, 15, 509-582.

[218] Taqqu, M.S., V. Teverovsky and W. Willinger (1995). Estimators for long-range dependence: An empirical study. *Fractals,* 3, 785-798.

[219] Taqqu, M.S., V. Teverovsky and W. Willinger (1997). Is network traffic self-similar or multifractal. *Fractals*, 5, 63-73.

[220] Taylor, C.C. and S.J. Taylor (1991). Estimating the dimension of a fractal. *J. Roayal Statistical Society B.* 53, 353-364.

[221] Tewfik, A.H. and M. Kim (1992). Correlation structure of the discrete wavelet coefficients of fractional Brownian motion. *IEEE transactions on information theory*, 38, 2, 2, 904-909.

[222] Tkacz, G. (2002). Estimating the fractional order of integration of interest rates using a wavelet OLS estimator. Technical report 2000-5, Department of Monetary and Financial Analysis, Bank of Canada.

[223] Tse, Y.K., V.V. Ahn and Q. Tieng (2002). Maximum likelihood estimation of the fractional differencing parameter in an ARFIMA model using wavelets. *Mathematics and Computers in Simulation*, 59, 153-161.

[224] Tsybakov, B. and N.D. Georganas (1997). On the self-similar traffic in ATM queues: definitions, overflow probablility bound, and cell delay distribution. *IEEE/ACM Trans*, Netw. 5. 3, 397-409.

[225] Velasco, C. (1999a). Non-stationary log periodogram regression. *J. Econometrics*, 91, 325-371.

[226] Velasco, C. (1999b). Gaussian semiparametric estimation of non-stationary time series. *J. time series analysis*, 20, 1, 87-127.

[227] Velasco, C. and P.M. Robinson (2000). Whittle pseudo-maximum likelihood estmation for nonstationary time series. *J. the American Statistical Association*, 95, 452, 1229-1243.

[228] Vetterli, M. and J. Kovačević (1995). Wavelets and subband coding. Prentice Hall, Englewood Cliffs, NJ.

[229] Vidakovic, B. (1998). Nonlinear wavelet shrinkage with Bayes rules and Bayes factors. *J. the American Statistical Association*, 93, 173-179.

[230] Vidakovic, B. (1999). Statistical modeling by wavelets. New York: Wiley.

[231] Vidakovic, B. and P. Müller(1999). An introduction to wavelets. In: Bayesian Inference in Wavelet-Based Models. EditorsMüller and Vidakovic, Springer-Verlag, Lecture Notes in Statistics, 141, 1-18.

[232] Veitch, D. and P. Abry (1999). A wavelet-based joint estimator of the parameters of long-rangedependence. *IEEE Transactions on Information Theory*, 45, 3, 878-897.

[233] Wang, R. (1998). Some properties of sums of independent random sets. *Northeastern Math. Journal*, 14(2), 203-210.

[234] Wang, R. and Z. Wang (1997). Set-Valued Stationary Processes. *J. Multivariate Analysis*, 63, 1, 180-198.

[235] Whitcher, B. and M.J. Jensen (2000). Wavelet estimation of a local long memory parameter. *Exploration Geophysics*, 31, 94-103.

[236] Whitcher, B. (2001). Simulating Gaussian stationary processes with unbounded spectra. *J. Computational and Graphical Statistics*, 10, 1, 112-134.

[237] Whitcher, B. (2004). Wavelet-based estimation procedures for seasonal long memory models. *Technometrics*, 46, 2, 225-238(14).

[238] Whittle, P. (1951) Hypothesis testing in time series analysis. Hafnerr, New York.

[239] Wickerhauser, M.V. (1996). Adapted Wavelet Analysis from Theory to Software. A. K. Peters, Wellesley, MA.

[240] Woodward, W.A., Q.C. Cheng and H.L. Gray (1998). A k-factor GARMA long-memory model. *J. Time Series Analysis*, 19, 5, 485-504.

[241] Wornell, G.W. (1996). Signal processing with fractals: a wavelet based approach. Prentice Hall, Englewood Cliffs, New Jersey.

[242] Yajima, Y. (1985). On estimation of long-memory time series models. *Australian and New Zealand Journal of Statistics*, 27, 3, 303-320.

[243] Yang, M. (2000). Some properties of vector autoregressive processes with Markov-Switching coefficients. *Econometric Theory*, 16, 23-43.